

## THESIS / THÈSE

### MASTER EN SCIENCES BIOLOGIQUES

**Exploration systématique de domaines d'interaction minimaux entre protéines essentielles à l'embryogénèse de *Caenorhabditis elegans*  
expérience pilote au moyen du "fragmentome" de 44 ORFS**

Devos, Maxime

*Award date:*  
2005

*Awarding institution:*  
Université de Namur

[Link to publication](#)

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



---

**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX  
NAMUR**

**Faculté des Sciences**

**EXPLORATION SYSTEMATIQUE DE DOMAINES D'INTERACTION MINIMAUX ENTRE  
PROTEINES ESSENTIELLES A L'EMBRYOGENESE DE *CAENORHABDITIS ELEGANS* :  
EXPERIENCE PILOTE AU MOYEN DU "FRAGMENTOME" DE 44 ORFS**

**Mémoire présenté pour l'obtention du grade de  
licencié en Sciences biologiques**

Maxime DEVOS  
Juin 2005

**Exploration systématique de domaines d'interaction minimaux entre protéines  
essentielles à l'embryogenèse de *Caenorhabditis elegans* : expérience  
pilote au moyen du "fragmentome" de 44 ORFs**

DEVOS Maxime

Résumé

Au cours de l'embryogenèse, les divisions cellulaires asymétriques et l'établissement de la polarité cellulaire sont des étapes critiques pour l'apparition de la diversité cellulaire. Se plaçant au point de vue protéique, un réseau d'interactions protéiques (ou interactome) au sein de *Caenorhabditis elegans* a été établi. L'hypothèse posée est que l'analyse de l'espace de cette carte interactomique, comportant les 802 ORFs essentielles à l'embryogenèse, recèle la clé du déroulement de celle-ci. L'élaboration d'une carte moléculaire de domaines d'interaction permettrait probablement d'interpréter fonctionnellement un tel réseau en termes de plasticité et de dynamique en rapport avec le développement.

Dans cette optique, nous avons procédé à un crible double-hybride (2H) utilisant à l'échelle pilote une librairie de fragments protéiques constituée à partir de 44 ORFs (environ 40 fragments par ORF) pour identifier les domaines d'interaction de sept partenaires de ces 44 ORFs avérés en test 2H "full length". Pour ce faire, nous avons testé chacun des sept interactants "full length" en appât contre la banque de fragments d'ORFs, utilisée comme proie.

Sur les sept protéines appâts testées contre la banque de fragments, seule la protéine rfc-4 a permis de mettre en évidence une interaction 2H. Les fragments pêchés avec cette protéine correspondent uniquement à ceux de sa protéine avérée, rfc-3, en "full length". Cette observation argue en faveur de la spécificité du test. La prédiction de domaines n'a pu être faite puisque les fragments de rfc-3 n'ont pas donné lieu à une interaction. Ces résultats feront l'objet d'une discussion.

Mémoire de licence en Sciences biologiques

Juin 2005

**Promoteur:** Pr. J. Vandenhoute

**Co-Promoteur:** Pr. M. Vidal

Au terme de ce travail, je tiens à remercier le Professeur J. Vandenhoute, promoteur de ce mémoire, ainsi que le Professeur M. Vidal pour m'avoir permis d'effectuer ce mémoire aux USA. Merci particulièrement pour votre aide et vos conseils si précieux, Professeur Vandenhoute.

Merci à Mike Boxem qui m'a encadré pendant ce séjour, ainsi qu'à tous les autres membres du Vidal lab. Je pense tout particulièrement à Tracey et Aby qui ont rendu le séjour si facile. Un immense merci à Nono pour le Colorado. Merci également à Joram pour m'avoir aidé durant les premières semaines. Les pizzas du vendredi me manquent...

Je remercie tout autant les membres de l'URBM pour leur disponibilité et leurs conseils.

Je tiens aussi à remercier les autres mémorants du labo et tout particulièrement Géraldine ("GG") sans qui ce voyage à Boston aurait été beaucoup moins drôle. Bravo à toi pour m'avoir supporté pendant trois mois, c'est un exploit !

Je n'oublie pas mes amis et mes chimistes adorées : Céline, les deux Isa, Cindy, Audrey ... avec qui ces années d'études ont été un vrai plaisir. Le groupe AB vaincra ! Merci aussi à Christina, et surtout, reste comme tu es, no matter what they say !

Merci à ma famille et tout particulièrement à mes parents pour m'avoir permis de réaliser ces études, tout ceci, n'aurait pu être réalisé sans eux. Je pense aussi à mes deux sœurs que j'adore!

Et finalement, je vous remercie, vous lecteurs et membres du jury, qui allez prêter attention à ce travail.

# Table des matières

<i>Liste des abréviations :</i>	<i>1</i>
---------------------------------	----------

## INTRODUCTION

<b>1. Le domaine protéique est un module</b>	<b>4</b>
<b>2. De l'interactome protéique à l'interactome de domaines</b>	<b>6</b>
2.1 L'interactome protéine-protéine	6
2.2 L'interactome domaine-domaine	6
<b>3. Approche expérimentale de l'interactome domaine-domaine</b>	<b>9</b>
3.1 Expériences "proof of principle"	9
3.2 Description de la stratégie de recherche de domaines d'interaction minimaux par "fragmentome"	10
<b>4. Cadre général du projet et objectif de ce mémoire</b>	<b>12</b>

## RESULTATS

<b>1. Génération de l'ORFéome de gènes essentiels aux premières étapes de l'embryogenèse de <i>C. elegans</i></b>	<b>13</b>
1.1. Etablissement de la liste des ORFs essentielles	13
1.2. Synthèse d'un amplicon de chaque ORF	13
<b>2. Vérification de la séquence des ORFs de l'ORFéome</b>	<b>15</b>
2.1. Transformation bactérienne	15
2.2. Résultats du séquençage	16
<b>3. Génération du "fragmentome"</b>	<b>17</b>
3.1. Position et longueur des amorces	17
3.2. Génération du "fragmentome" pour 44 ORFs	18
<b>4. Clonage et mise en "pool" des fragments</b>	<b>19</b>
<b>5. Utilisation du "fragmentome 44" pour la détection de domaines d'interaction</b>	<b>20</b>

## DISCUSSION

<b>1. Discussion</b>	<b>24</b>
1.1 Le nombre d'interactions du test 2H avec le "fragmentome" est faible	24
1.2 Le test n'a été positif que pour le fragment "full length" de rfc-3	26
<b>2. Perspectives</b>	<b>30</b>
<b>3. Conclusions</b>	<b>31</b>

## MATERIEL ET METHODES

<b>1. Restriction/ligation d'ADN</b>	<b>32</b>
1.1. Enzyme de restriction	32
1.2. Ligation d'un insert dans un vecteur :	32
1.3. Utilisation des enzymes de restriction/ligation pour cloner les fragments:	32
<b>2. Production, purification et analyse de produits PCR</b>	<b>33</b>
2.1. Réaction PCR	33
2.2. Purification des produits PCR via le kit QIAquick :	33
2.3. Electrophorèse en gel d'agarose	34
<b>3. Préparation d'ADN bactérien</b>	<b>34</b>
3.1. Minipréparation avec le kit QIAprep Spin Miniprep	34
3.2. Minipréparation avec plaques à 96 puits profonds	34
3.3. Maxipréparation avec le kit Hispeed Plasmid Midi	34
<b>4. Milieux de culture utilisés</b>	<b>34</b>
4.1. Milieu bactérien	34
4.2. Milieu pour levures	35
<b>5. Transformation</b>	<b>36</b>
5.1. Transformation bactérienne	36
5.2. Transformation de levures	37
<b>6. Crible double-hybride (2H) :</b>	<b>38</b>
6.1. Principe	38
6.2. Le système 2H par co-transformation	38
6.3. Phénotypage des colonies :	39
6.4. PCR sur les colonies de levures :	40
<b>7. Souches utilisées</b>	<b>41</b>
<b>8. Réactions de MultiSite Gateway™</b>	<b>41</b>
8.1. Réaction BP Clonage	41
8.2. Réaction LR Clonage	41

## BIBLIOGRAPHIE

## Liste des abréviations :

µg	Microgramme
µL	Microlitre
2H	Double-hybride
3AT	3-amino-triazole
A <sup>600</sup>	Absorbance à 600 nm
AD	Domaine de transactivation de gal4 (activation domain)
ADN (DNA)	Acide désoxyribonucléique
ADNc	ADN complémentaire
Amp	Ampicilline
attB	Site d'attachement de la bactérie
attP	Site d'attachement du phage
BSA	Albumine de sérum bovin
<i>C. elegans</i>	<i>Caenorhabditis elegans</i>
DB	Domaine de liaison à l'ADN de gal4
ddH <sub>2</sub> O	Eau distillée et désionisée
dNTP	Déoxynucléotide triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	Acide éthylènediamine tétraacétique
HCl	Acide chloridrique
HIF	Facteur d'intégration de l'hôte (Host Integration factor)
His	Histidine
IDA	Interaction Defective Allele
Int	Intégrase
kb	kilobase
LB	Milieu bactérien Luria Bertani
Leu	Leucine
LiAc	Acétate de lithium
réaction LR	Réaction "gauche-droite" (Left-Right Reaction)
ORF	Phase ouverte de lecture (Open Reading Frame). Séquence d'ADN entre le codon start et le stop
pb	Paire de bases
PCR	Réaction de polymérisation en chaîne (Polymerase Chain Reaction)
PEG	Polyéthylène glycol
Pfam	Protein FAMily database
réaction BP	Réaction bactérie-phage (Bacteria-Phage reaction)
rfc	Replication Factor c
RNAi = ARNi	ARN interférence
RPM	Rotations par minute
RT	Transcriptase inverse (reverse transcriptase)
S.O.C.	Milieu "Salt Optimized Carbon"
SC	Milieu synthétique complet (Synthetic Complete medium)
TBE	Tris Borate EDTA
TE	Tris EDTA
Trp	Tryptophane

Ura	Uracile
WFMP	Worm Fragmentome minipep
WFPCR	Worm Fragmentome PCR
X-gal	5-bromo-5-chloro-3-indolyl- $\beta$ -Dgalactoside
Xis	Excisionase
YEPD	Milieu "Yeast extract peptone dextrose"
$\beta$ -gal	$\beta$ -galactosidase



## **Avant-propos :**

Le progrès dans l'accumulation des informations en biologie moléculaire suit deux voies principales : l'obtention de séquences et, plus lentement celle de structures protéiques. Ces répertoires, même si on les imagine complets ne donneraient pas pour autant la clé au problème de la compréhension des mécanismes biologiques. En effet, la relation structure-fonction n'est pas déductible, comme on a voulu le croire, de la seule connaissance des modèles cristallographiques d'une protéine donnée. C'est que, en réalité, la relation gène-enzyme-fonction n'est pas univoque et que la définition même du gène comme unité de fonction n'est plus généralement tenable. Les fonctions, ou mieux, les rôles joués par les protéines ne le sont que dans un contexte donné. D'une certaine manière, tout peptide est une sous-unité, au moins fonctionnelle d'un ensemble complexe d'autres partenaires. Ainsi, une des découvertes récentes de la biologie moléculaire et cellulaire est sans conteste l'existence dans la cellule de "machineries" multicomposites dédiées localement à l'exercice de fonctions complexes. Alors que certaines de ces machineries sont relativement figées une fois construites (auto-assemblage) d'autres demeurent dans un état permanent de fluctuation (auto-organisation) dynamique ("turn over" de composants ou modifications de leur composition même) (Misteli 2001). En effet, si on considère les machineries parmi les mieux connues de la cellule, telles le ribosome ou, le spliceosome ou encore le transcriptosome on réalise clairement que la complexité de leur fonctionnement réside dans la multiplicité de leurs composants mais aussi dans le fait que des changements importants peuvent survenir dans cette composition au plan quantitatif (arrivée de nouveaux composants) et surtout qualitatif (modification chimique covalente ou non d'un ou plusieurs de ceux-ci). On sait par exemple que selon l'étape du cycle traductionnel, le ribosome voit sa masse pratiquement doubler. Même le nucléosome dont la fonction semblait être l'assemblage en un échafaudage fixe avec la macromolécule la plus définie et constante qui soit, l'ADN, subit lui-même non seulement des modifications d'ordre post-traductionnel mais aussi des changements de composition (Redon et al. 2002). La clé des mécanismes moléculaires sous-tendant ces processus d'autogenèse régulés est à rechercher dans les reconnaissances cognitives ou entre les co-aptations paires de ces composants.

L'établissement de la séquence génomique des organismes et l'émergence de techniques à haut débit capables d'investiguer divers aspects fonctionnels d'une protéine (localisation, interaction...) fonde ce qu'on a coutume de désigner du vocable de post-génomique ou mieux de "physio-génomique". La mise en œuvre par des groupes (Li et al. 2004 ; Fromont-Racine, Rain et Legrain 2002) d'analyses double-hybride (2H) à haut débit a fourni des cartes d'interaction physique correspondant aux protéomes de divers organismes tels que la levure, le ver, la drosophile et bientôt l'homme (Vidal, sous presse).

Notre projet est de fragmenter des protéines partenaires au sein de l'interactome du ver *C. elegans*, ou du moins d'un sous-ensemble ou "espace" de celui-ci, afin de soumettre ces fragments clonés ("fragmentome") aux tests d'interaction 2H. L'hypothèse qui est faite est que cette méthodologie est susceptible de révéler des fragments peptidiques constituant des domaines d'interaction spécifiques entre protéines partenaires. C'est donc à une dissection fonctionnelle des éléments ou domaines d'interaction protéine-protéine que nous visons dans le but de décrypter les mécanismes moléculaires assurant les associations structurales, statiques ou dynamiques des composants des machineries protéiques au sein des organismes.

L'introduction qui suit va détailler le descriptif global qui vient d'être fait de la problématique du travail et va préciser l'approche et l'objectif concret que nous lui assignons.

## **1. Le domaine protéique est un module**

Si on entend par module une entité structurale et fonctionnelle dotée d'une certaine autonomie (Fraser et al. 2005) on voit que les groupes de protéines interconnectées d'un sous-réseau dont les interacteurs conditionnent un ou plusieurs rôles au sein du système vivant répondent à cette définition. A une échelle plus réduite, on peut considérer que, au niveau de la protéine elle-même, on a affaire à une structure modulaire. En effet, la protéine est analysable en termes d'entités structurales et fonctionnelles. Ainsi une succession séquentielle d'acides aminés peut se constituer en une structure peptidique secondaire à rôle défini ; il peut s'agir d'un domaine participant à un centre actif, à une surface d'interaction intrapeptidique, interpeptidique ou encore d'un domaine assurant la liaison avec un ligand ou substrat. Certains de ces domaines d'interactions répondent à des types de structures présentes parfois en centaines de copies dans le protéome humain, particulièrement dans les modules protéiques chargés des voies de signalisations. C'est ainsi que ces séquences peuvent être souvent identifiées grâce à leur consensus acides aminés, permettant ainsi de prédire les propriétés de liaison et les fonctions biologiques d'une protéine (Copley, 2002). La cellule semble utiliser un lot limité de ces domaines d'interaction, un peu comme des blocs de construction qui sont assemblés selon différentes combinaisons dans les protéines (Figure 1) (Pawson et Nash 2003). A ces domaines dits séquentiels s'ajoutent des domaines structurés cette fois grâce au "folding" tridimensionnel associant des acides aminés non alignés et même souvent très distants dans la structure primaire mais dont la disposition dans l'espace contribue elle aussi à une fonction.

Ces considérations très simples qui émergent des connaissances de la biologie moléculaire structurale actuelle renvoient à un débat ancien qui est celui de la définition du gène ou plutôt des définitions du gène. Rappelons que le concept de "cistron" permet d'assigner au niveau d'un locus génétique des bornes ou limites à un gène singulier : toute mutation qui affecte la fonction considérée est à localiser dans le gène ou cistron. Cette définition purement fonctionnelle complétée par la conception d'une relation univoque entre le gène, morceau d'ADN, et la protéine qu'il code va avoir un impact exceptionnel sur ce qui va devenir au cours du dernier demi-siècle, la génétique moléculaire. Cependant certaines complexités déjà perçues par la génétique classique, telles que par exemple l'hérédité polygénique ou la pléiotropie, ne rentraient pas de manière satisfaisante dans le cadre de la définition d'un gène assimilé à une séquence d'ADN. Ces dernières années, une nouvelle vision de la fonction des gènes est apparue : les produits des gènes fonctionnent sous forme de voies et de réseaux très complexes utilisés, en tout ou en partie, pour accomplir des tâches différentes. Par conséquent, chaque produit de gène participe à la réalisation de multiples caractères (action dite pléiotropique des gènes), et chaque caractère résulte de l'action conjuguée de nombreuses protéines (origine multigénique des caractères). De plus, ces réseaux de protéines varient, c.-à-d. que leur structure décrite par les interactions entre leurs composants se modifie, en fonction de paramètres comme les conditions environnementales (milieu extracellulaire, type cellulaire...), le "background" génétique ou encore les conditions temporelles (phase de développement...).

Une clé dans la compréhension de l'action intégrée de l'ensemble des produits d'expression du génome, essentiellement les protéines (ou protéome) pour déterminer les phénotypes (ou phénomène) repose sur la description instantanée des interactions s'exerçant entre les acteurs (ou nœuds) d'un réseau. Un sous-réseau d'un type peut s'adapter et évoluer en un sous-réseau d'un autre type structural et fonctionnel en modifiant simplement les interactions entre ses composants (Figure 2). Dans une telle représentation on voit

l'importance du sous-réseau, ou module, comme une entité fonctionnelle (semi-) autonome dans l'ensemble des protéines en réseau d'une cellule. En outre, on voit clairement aussi qu'un acteur (par exemple la protéine ou le nœud central) ne joue pas seulement un rôle mais au contraire plusieurs, illustrés par les diverses connections qu'il entretient avec les autres acteurs du module. La protéine est donc elle-même analysable en modules distincts (représentés par autant de petits cercles de couleur dans un noeud). Le principe "one gene - one protein - one function" ne tient donc plus.

Derrière ces approches, on retrouve aussi manifestement une remise en cause d'une certaine forme de réductionnisme qui tentait précisément de réduire les propriétés complexes des organismes vivants aux caractéristiques d'une ou de quelques macromolécules biologiques, c.-à-d. d'un ou de quelques gènes ("one-gene-at-a-time"). Les approches "post-génomiques" visent à chercher l'ordre du vivant, et donc son explication, en analysant ses composants macromoléculaires non de façon singulière, mais selon une approche synthétique pour comprendre la manière dont ces composants s'assemblent et "fonctionnent" ensemble. Cependant, les analyses à haut débit et les études à l'échelle d'un ou de quelques gènes sont deux méthodes qui se complémentent. En effet, les études fonctionnelles à l'échelle du génome permettent l'élaboration de nouvelles hypothèses biologiques qui pourront ensuite être confirmées et interprétées en détails par des analyses plus réductionnistes. Prenons par exemple le cas d'une protéine enzymatique, Arg 5,6. Celle-ci est connue de longue date pour intervenir dans le métabolisme de l'ornithine. Etonnamment, un test haut débit de CHIP a pêché Arg 5,6 ce qui suggère qu'elle se liait à l'ADN. Des études complémentaires ont confirmé que cette protéine était bien une entité multifonctionnelle qui pouvait, en plus de son rôle métabolique, agir comme facteur de transcription de gènes mitochondriaux et nucléaires (Hall et al. 2004; Bhardwaj et al. 2005). Dans un tel cas, on voit la valeur heuristique de cette approche puisque le test à haut débit, sans a priori ou hypothèse de travail, a permis d'identifier une nouvelle fonction qui autrement serait probablement restée ignorée.

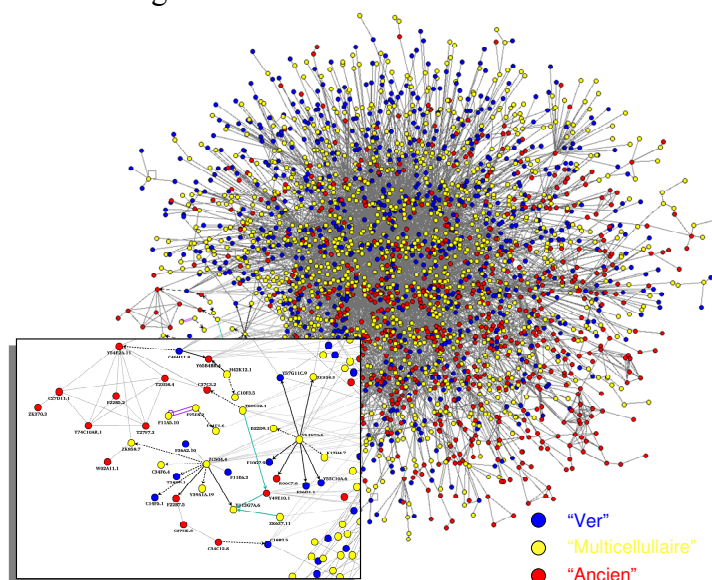
En résumé, conceptuellement l'organisme vivant est à considérer sous l'angle de ses acteurs moléculaires, en particulier les protéines, établissant un vaste réseau d'interacteurs. Les sous-réseaux de la carte interactomique en sont des modules fonctionnels et les nœuds eux-mêmes peuvent être analysés comme des associations de modules, les domaines. C'est ce que Hartwell a appelé le passage de la biologie moléculaire à la biologie modulaire (Hartwell et al. 1999)

## 2. De l'interactome protéique à l'interactome de domaines

### 2.1 L'interactome protéine-protéine

Comme on vient de le voir, il est important de considérer la fonction d'un gène dans le contexte d'un réseau d'interactions moléculaires. Pour ce faire, une cartographie extensive des interactions protéine-protéine à l'échelle d'un organisme a été entreprise, c'est l'interactome (Li et al. 2004) (Figure 3). En pratique, un graphe est dessiné dans lequel les gènes et leurs produits sont représentés par des nœuds, ou "nodes", et les interactions sont représentées par des liens, ou "edges". A ce jour chez le nématode *C. elegans* par exemple, l'interactome d'environ 3000 protéines comporte quelque 5000 interactions protéine-protéine potentielles.

L'objectif à terme de réaliser un "atlas" composé de cartes moléculaires obtenues à partir de l'intégration de cet interactome avec des bases de données d'expression génique et des criblages phénotypiques par ARNi à l'échelle du génome (Walhout et al. 2002). Le fait de disposer de données de protéomique à grande échelle nous aide donc à accéder à un niveau de compréhension globale de l'organisme.



**Figure 3.** Graphe représentant l'interactome de *C. elegans* dans lequel les gènes et leurs produits sont représentés par des nœuds, ou "nodes", et les interactions des trois types décrits sont représentées par des liens, ou "edges". Les nœuds sont colorés suivant leur conservation évolutive : en rouge, les 784 protéines présentant clairement un orthologue chez la levure ("ancien") ; en jaune, les 1314 protéines possédant un orthologue chez la drosophile, *Arabidopsis*, ou l'humain mais pas chez la levure ("multicellulaire") et en bleu, les 836 protéines sans aucun orthologue en dehors de *C. elegans* ("ver"). Le cadre met en évidence une petite partie de ce réseau. (Li et al. 2004).

### 2.2 L'interactome domaine-domaine

L'organisation biologique repose sur une hiérarchie de niveaux structuraux, chacun des niveaux s'édifiant à partir du niveau inférieur (Oltvai et Barabási 2002) (Figure 4 et Figure 5). Les protéines forment la classe la plus nombreuse et la plus diversifiée des molécules biologiques; elles peuvent constituer plus de 50% du poids sec des cellules. Elles ont un rôle dans pratiquement tous les aspects structuraux et fonctionnels des cellules. Le niveau d'organisation inférieur est celui des domaines protéiques qui peuvent être considérés comme

les unités fonctionnelles, ou modules, des protéines capables d'une activité et d'une structuration ("folding") relativement autonome (Figure 6).

### **2.2.1 Avantages de l'interactome domaine-domaine**

Le type de carte qu'est l'interactome domaine-domaine représenterait un outil de qualité pour la prédiction et l'interprétation de voies de régulation cellulaire. Illustrons cette idée avec un cas bien connu : celui de la régulation du passage en phase G1 dans des cellules devant subir une mitose. Différentes protéines sont impliquées dans cette étape capitale dont les protéines E2F, Rb, HDAC ou DP1. La Figure 7 permet d'observer les interactions entre ces différentes protéines telles qu'on peut les voir dans l'interactome. Ce type de données d'interaction, ne dit pas, par exemple, si la protéine Rb peut lier E2F et HDAC simultanément ou non. En revanche, s'intéresser à un niveau plus fin d'organisation, celui des domaines, pourrait permettre de comprendre plus précisément la dynamique de ce petit réseau de protéines. L'observation que les domaines de E2F qui permettent l'interaction avec Rb, DP et l'ADN sont différents permet par exemple de postuler que ces quatre protéines pourraient former un complexe unique qui fixe l'ADN comme un tout. Sachant par ailleurs que E2F permet une activation de la transcription alors que, au contraire, HDAC l'inhibe et que Rb fixe le même domaine d'activation de la transcription de E2F, on peut comprendre que Rb fixe E2F et bloque l'activation de la transcription. Toutes ces informations sont déjà connues mais cet exemple démontre clairement que les informations apportées par la connaissance des domaines auront une valeur heuristique qui permettra d'arriver à des modèles biologiques détaillés applicables aux réseaux de protéines identifiés à l'échelle de l'interactome.

### **2.2.2 Stratégies utilisées pour générer l'interactome domaine-domaine**

En principe, et compte tenu de l'existence de domaines d'interaction conformationnels et séquentiels, l'interactome domaine-domaine doit recouvrir les deux types. En ce qui concerne les domaines non linéaires (ou conformationnels) d'une protéine ou d'un nœud donné, on peut imaginer une approche expérimentale qui consisterait à générer des mutants de cette protéine affectant des liens singuliers, ou "edges" (approche "edgétique"), tout en conservant toutes les autres interactions. On dira des allèles affectant un et un seul "edge" d'une protéine (IDA ou "interaction defective alleles") qu'ils sont de nature "edgétique". La localisation sur le modèle 3D de cette protéine de l'ensemble des mutations appartenant à un "groupe de disruption" (ou symétriquement "groupe d'interaction") de l'interaction pairée définirait le domaine d'interaction.

Par ailleurs, certaines tentatives sont entreprises pour tenter de prédire *in silico* les domaines d'interaction entre paires de partenaires d'une carte interactomique donnée. Le principe est de modéliser (le plus souvent par homologie) la structure des partenaires et ensuite d'éprouver *in silico* la vraisemblance d'une interaction entre eux. On parle parfois de 2H *in silico* (Pazos et al 2002). Les succès de cette approche sont encore rares à ce jour. Dans la même voie on pourrait imaginer qu'une fois qu'un certain nombre de paires homologues de domaines d'interaction (des "interlogues" de domaines) seront connus et caractérisées structuralement et fonctionnellement, la recherche par homologie via l'alignement de séquences pourrait éventuellement permettre de retrouver des partenaires potentiels au niveau de tout un génome.

L'étude de la structure tridimensionnelle des protéines par diffraction des rayons X sur des monocristaux de protéines (cristallographie) pourrait avoir la capacité de déterminer aussi

bien les domaines linéaires que conformationnels. Cependant, il n'est actuellement pas réaliste d'appliquer cette stratégie longue et ardue à l'ensemble du protéome cellulaire, d'autant que pour réellement mettre une interaction en évidence, il faut, à chaque fois, être capable de cristalliser les deux protéines partenaires dans un même complexe.

Une stratégie qui s'impose à l'esprit lorsqu'il est question d'identifier des interactions physiques entre protéines est évidemment la technique du 2H et ses multiples dérivés. En ce qui concerne en particulier les domaines séquentiels de protéine l'approche 2H qui fut à la base des cartes interactomiques pourrait être exploitée efficacement mais également de manière extensive (à haut débit). L'idée qui sera développée au concret plus loin consiste à générer pour un protéome donné les fragments peptidiques ("fragmentome"<sup>1</sup>), de taille compatible avec une représentation de domaines (de l'ordre de 200 acides aminés) et à les tester en paires dans un test 2H. La taille de la matrice complète d'interaction domaine-domaine à générer sera, comme nous le verrons, d'une dimension près de 50x supérieure à l'interactome protéine-protéine.

---

<sup>1</sup> Le "fragmentome" est pris ici dans son sens "extensif" c.-à-d. correspondant à l'ensemble de tous les fragments peptidiques d'un protéome. Cependant, dans la suite de ce travail, nous désignerons du terme de "fragmentome" l'ensemble des fragments d'une protéine ou de quelques protéines dans un espace donné de l'interactome.

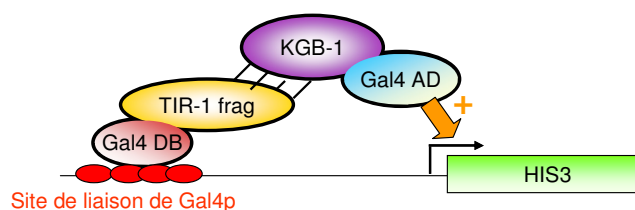
### 3. Approche expérimentale de l'interactome domaine-domaine

#### 3.1 Expériences "proof of principle"

##### 3.1.1 Définition du domaine d'interaction de TIR-1 avec KGB-1 par analyse des fragments de délétion

La technique utilisée dans le cadre de ce mémoire est une application à grande échelle d'une stratégie développée dans le laboratoire de Marc Vidal. Utilisée par Pierre-Olivier Vidalain et Stuart Milstein, elle permet de définir les régions d'une séquence protéique permettant une interaction physique avec une autre protéine. Pour ce faire, le cadre de lecture ouverte, ou ORF, du gène correspondant à la protéine étudiée est divisé en fragments divers de longueurs variées, afin d'exprimer les domaines fonctionnels individuels de la protéine. L'utilisation d'un "fragmentome" en appât dans un test double-hybride (2H) permet en principe à la fois de découvrir un maximum de partenaires et d'identifier directement le domaine minimal d'interaction de la protéine d'intérêt avec ses partenaires.

TIR-1 est la protéine sélectionnée pour subir ce processus de fragmentation. Elle est connue pour interagir avec la protéine KGB-1. Le but va être de définir la fraction de TIR-1 nécessaire à l'interaction avec KGB-1. Des réactions PCR sont réalisées pour générer les différents fragments de l'ORF. Avec des amorces disposées environ toutes les 75 paires de bases il est possible de générer plus de 100 fragments différents. Cette fragmentation systématique est effectuée suivant une matrice de délétion (Figure 8) Chaque fragment est cloné individuellement dans un vecteur 2H DB-X puis transformé en levure. L'interaction avec la protéine "full length" KGB-1, mise en appât, et la matrice de fragments est testée (Figure 9). Le fragment minimal de TIR-1 nécessaire pour l'interaction avec KGB-1 est déterminé, il est codé par les nucléotides 121 à 360 (Figure 10), région correspondant au domaine prédit comme la zone de faible complexité (Figure 11). Cette expérience constitue la preuve de principe de la possibilité d'identifier des interactions de domaines par 2H en levure.

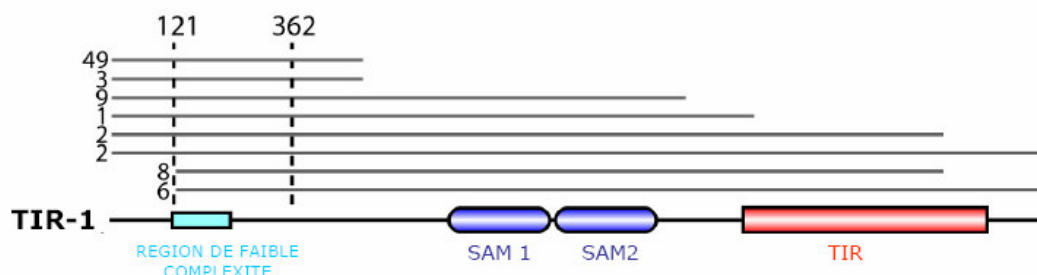


**Figure 9.** Schéma du test 2H "fragmentome". Chaque fragment (TIR-1 frag) est testé en double-hybride contre la protéine KGB-1 "full length". Si le fragment et la protéine KGB-1 interagissent physiquement (lignes obliques) le facteur de transcription Gal4 est recomposé et peut activer (flèche +) la transcription du gène rapporteur lacZ.

##### 3.1.2 Simulation d'une recherche de ce domaine d'interaction par un crible

Pour déterminer si il est possible d'identifier la même interaction à partir d'un pool de fragments, ce qui sera la condition expérimentale d'un "screen" grandeur réelle, les différents fragments PCR de TIR-1 sont "poolés". Les levures contenant le plasmide appât KGB-1 "full length" sont transformées avec la librairie de fragments de TIR-1 et étalées sur des plaques

sélectives. Les colonies positives sont isolées et l'identité du fragment de TIR-1 est déterminée par PCR et séquençage. Le fragment minimal requis pour l'interaction de TIR-1 avec KGB-1 correspond exactement à celui identifié dans l'expérience où chaque fragment était testé individuellement (Figure 12). Il est donc possible d'utiliser cette stratégie pour identifier, cette fois dans des conditions à haut débit, les domaines d'interaction de protéines d'un sous-réseau ou module donné.



**Figure 12.** Fragments de TIR-1 isolés depuis un pool de fragments en utilisant la protéine "full length" KGB-1 comme appât. Les fragments sont comparés à TIR-1 "full length". Les lignes discontinues indiquent le domaine minimal d'interaction commun à chaque fragment qui s'étend du nucléotide 121 à 362.

### **3.2 Description de la stratégie de recherche de domaines d'interaction minimaux par "fragmentome"**

Nous décidons d'identifier les domaines d'interaction de 721 gènes indispensables aux premières étapes de l'embryogenèse de *C. elegans*. Dans ce but chaque protéine, classifiée par ailleurs comme essentielle à l'embryogenèse, sera testée en 2H contre une librairie complexe de fragments, ou "fragmentome", générée à partir de ces mêmes protéines. Une approche PCR non biaisée sera utilisée pour générer cette librairie afin de s'assurer que toutes les interactions potentielles soient représentées, aussi bien les prédites que les inconnues. La première étape est d'obtenir un clone de la phase ouverte de lecture (ORF) prédite pour chacun des 721 gènes de l'embryogenèse. Pour chaque ORF amplifiée et clonée avec succès, des amorces "forward" et "reverse" seront conçues à environ 200 pb d'intervalle et toutes les combinaisons possibles d'amorces serviront à générer des fragments PCR des ORFs. Cela aboutit à un nombre moyen de 40 fragments par ORF. Les amplicons PCR seront collectés dans des "pools" d'environ 1000 fragments de même taille et transférés dans un vecteur 2H AD-Y approprié. Ces "pools" de constructions 2H seront ensuite mélangés pour générer la librairie de fragments finale.

Pour cribler la librairie de fragments avec les différents gènes une approche similaire au traditionnel crible d'ADNc sera utilisée. Le système 2H utilisé a été optimisé pour minimiser l'identification de faux positifs (cfr matériel et méthodes). Les cellules de levure exprimant la protéine appât "full length" sont transformées avec la librairie de fragments. Les interactions positives seront identifiées en étalant les levures transformées sur un milieu sélectif. L'identité des fragments interagissant sera déterminée par PCR et séquençage.

Dans les cribles 2H, tout comme dans d'autres techniques de haut débit, certaines interactions ne sont pas décelées et constituent des "faux négatifs" qu'on souhaite réduire au maximum. Nous traiterons ici les avantages que possède une librairie de fragments telle que celle que nous allons générer dans ce mémoire par rapport aux traditionnelles banques



d'ADNc et d'ORF. Quand on utilise une banque d'ADNc, l'inconvénient est que la librairie n'est pas normalisée et que dès lors certains clones sont surreprésentés par rapport à d'autres. En revanche, ce type de banque contient de nombreux clones qui ne sont pas "full length" par le fait que la RT avorte souvent en cours d'opération sur ses matrices d'ARNm. Or une protéine de fusion, appât ou proie, ou un fragment protéique peuvent ne pas avoir la même capacité d'interaction dans un test 2H. En effet, il est maintenant bien connu qu'un certain nombre de constructions dont nombre de constructions "full length", ne donnent pas des protéines de fusion (GAL4 DB ou AD + ADNc) capables d'interagir en 2H. Quelquefois, l'ajout ou le retrait de quelques acides aminés à la protéine en fusion est suffisant pour rendre la fusion fonctionnelle en 2H. Ce genre d'observation conduit à la conclusion qu'une interaction 2H peut apparaître ou disparaître suite à une modification de longueur (même minime) des peptides ou fragments de peptides en jeu. Dès lors une interaction 2H peut se manifester ou non selon que l'environnement natif soit présent ou non dans le test. Celui-ci reconstituant imparfaitement cet environnement il se peut que, comme évoqué plus haut, dans les conditions du test, un fragment d'une protéine exhibe l'interaction mieux, ou au contraire, moins bien que la protéine complète. A ce stade rien ne permet de décider si l'interaction ou l'absence d'interaction sont ou non des faux-positifs, ou faux-négatifs, respectivement. Un candidat 2H doit, en toute hypothèse, être éprouvé et validé par d'autres expériences. De fait, certains fragments correspondant à des domaines séquentiels d'interaction réelle *in vivo* peuvent ne plus n'interagir dans une fusion comprenant la protéine complète probablement parce la structure en fusion masque ce fragment. A contrario, des fragments non réactifs *in vivo* peuvent donner lieu à des faux positifs dans le test 2H. On comprend que l'utilisation d'une banque d'ORFs "full length" (ORFéome) possède des avantages et inconvénients en partie opposés à une banque d'ADNc. La librairie ORF est normalisée mais si l'ORF "full length" ne permet pas la formation d'une protéine de fusion fonctionnelle, l'interaction native ne sera jamais retrouvée. L'espoir est que notre librairie de fragments systématique ("fragmentome") va combiner les forces des deux librairies : la normalisation des séquences et la présence de fragments d'ORFs de longueur différente susceptibles de dévoiler l'ensemble des domaines potentiels séquentiels d'interaction.

Les résultats d'un essai conduit précédemment vont illustrer les potentialités de l'approche "fragmentome". On a observé effectivement dans ce cas que l'utilisation d'une librairie de fragments a permis une plus grande couverture que les protéines "full length". Comme le montre le test d'interaction ci-dessous impliquant 10 protéines impliquées dans le positionnement du fuseau mitotique, les 6 fragments incluant le fragment "full length" ont été générés pour chaque protéine et testés en matrice 2H. En utilisant 6 fragments par protéine, un nombre total de 13 interactions a été identifié (Figure 13). Par contre si seuls les fragments "full length" sont utilisés, seulement 8 interactions sont retrouvées. Une de ces interactions, uniquement retrouvée avec les fragments, a déjà été décrite ailleurs, confirmant la spécificité du test. On peut donc conclure que l'utilisation de fragments, plutôt que de protéines "full length", permet d'augmenter la couverture du test 2H.

## **4. Cadre général du projet et objectif de ce mémoire**

Générer des cartes d'interaction domaine-domaine pour l'intégralité des protéines de l'interactome représente un travail énorme. C'est pourquoi, nous avons décidé de nous focaliser sur un des sous-réseaux, ou espace réduit, de l'interactome : celui des protéines essentielles aux premières étapes de l'embryogenèse du ver nématode *C. elegans*.

Notre choix a été motivé par le fait que l'embryogenèse de *C. elegans* manifeste clairement une caractéristique fondamentale à la base même du développement des organismes. En effet, on observe chez ce ver une première division hautement asymétrique de l'embryon (Figure 14) (Lyczak et al. 2002). Ce processus génère des cellules-filles au devenir différent et peut dès lors être considéré comme une première étape visible de différenciation cellulaire. Cette division asymétrique requiert l'établissement d'un axe de polarité, suivi par une ségrégation asymétrique de constituants cellulaires et d'une orientation du faisceau mitotique afin de déterminer le plan de clivage. Des défauts dans l'établissement de cette polarité mènent à un développement anormal (Vallee et al. 2001) et à l'apparition de tumeurs (Bissell et al. 2001). Des approches génétiques ont déjà permis d'identifier de multiples gènes impliqués dans ce processus. Cependant, une compréhension globale des voies de transduction amenant à une cette division asymétrique est toujours manquante. Ce projet propose d'utiliser une approche originale, protéomique et fonctionnelle, pour identifier les liens entre les composants impliqués dans la polarité cellulaire et la division cellulaire asymétrique de *C. elegans*.

Le but ultime de ce projet est de générer et analyser un réseau d'interaction pour ces 721 gènes indispensables aux premières étapes de l'embryogenèse de *C. elegans*. Pour ce faire, les données d'interactions définies dans un crible 2H seront comparées et combinées aux caractérisations phénotypiques [résultant de mesures de coefficients de corrélation de Pearson à partir d'une série de 47 phénotypes testés systématiquement dans des expériences d'interférence ARN] et aux données d'expression [résultat de mesures des coefficients de corrélation de Pearson à partir de données d'expression obtenues par puces à ADN dans de nombreuses conditions expérimentales] pour ces 721 protéines. Un atlas de cartes en réseaux sera alors généré dans lequel les protéines seront connectées sur base d'interactions physiques, de similarité de profil d'expression et de similarité phénotypique. Pour chacune des interactions, il sera possible de déterminer les domaines d'interaction impliqués (Figure 15).

L'objectif concret de ce mémoire est de préparer de manière efficace le test d'interaction qui sera effectué sur l'ensemble des protéines essentielles à l'embryogenèse de *C. elegans*. En d'autres termes, définir la stratégie la plus appropriée pour convertir l'approche systématique du "fragmentome", telle qu'utilisée en bas débit dans l'expérience "proof of principle" décrite, en une approche utilisable dans des analyses à plus grande échelle ("High-Throughput Fragmentome"). Dans un premier temps, une collection d'ORFs (ORFéome) de séquences "wild-type" est à confectionner pour générer ensuite, par réaction PCR, les fragments des différentes protéines. Une mini-banque de fragments pour 44 des 802 ORFs sera utilisée dans le cadre d'un test pilote reconstituant dans une certaine mesure les conditions du test final à grande échelle. Son objectif est de déterminer les domaines d'interaction entre protéines pour lesquelles une interaction 2H en "full length" a déjà été montrée.

# **1. Génération de l'ORFéome de gènes essentiels aux premières étapes de l'embryogenèse de *C. elegans***

## **1.1. Etablissement de la liste des ORFs essentielles**

Une récente étude d'ARN interférence à l'échelle du génome a identifié 662 gènes essentiels aux premières étapes de l'embryogenèse de *C. elegans* (Hyman, communication personnelle). Un réseau connectant toutes ces protéines, ou "interactome", a été établi avec comme critère que chaque paire d'interacteurs retenue répond à au moins deux évidences expérimentales parmi les suivantes : phénotype similaire de mutants ARNi (Kamath et al. 2003 ; Fernandez et al. 2005), même profil d'expression sur DNA microarrays (Reinke et al. 2000), interaction physique en double-hybride (Li et al. 2004). La liste de 662 gènes indispensables au bon déroulement des étapes précoces de l'embryogenèse a été étendue à 721 gènes sur base des données non publiées du groupe de Kris Gonsalus (Gonsalus, communication personnelle). Tenant compte des variants d'épissage alternatif et en éliminant les ORFs dont la taille est supérieure à 5 kb, un nombre total de 802 ORFs a été retenu. Le premier objectif étant de tester les protéines correspondantes dans un test double-hybride (2H) en levure, il s'agit donc que ces clones "full length" soient dans des plasmides compatibles avec ce type de test.

Bien que dans les 802 ORFs à cloner, 537 soient déjà présentes dans la ressource ORFéome de *C. elegans* (Lamesch et al. 2004) et donc directement testables en 2H, l'arrivée sur le marché d'une nouvelle polymérase (KOD polymérase, cfr. matériel et méthodes) possédant un taux de fidélité 10x supérieur à l'enzyme utilisée pour générer l'ORFéome nous a poussé à recloner la totalité des 802 ORFs. En effet, comme dit plus haut nous avons décidé d'utiliser comme "template" pour générer le "fragmentome" une séquence de départ identique à celle du génome de référence (considéré comme "wild-type"). L'utilisation de la KOD polymérase nous permettra d'éviter au maximum la création de mutants PCR et garantira de la sorte une fraction plus importante d'amplicons "wild-type".

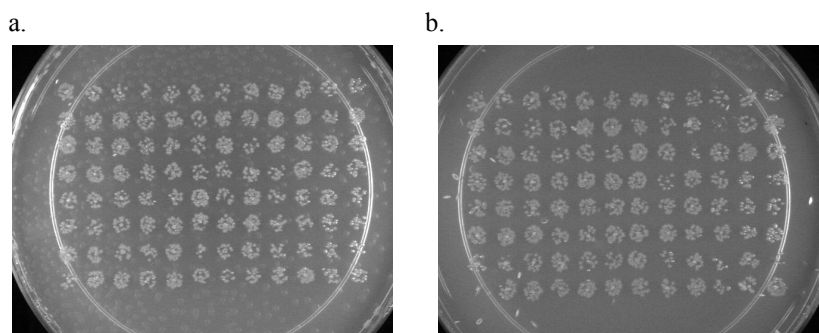
## **1.2. Synthèse d'un amplicon de chaque ORF**

Des amorces "forward" et "reverse" pour les 802 ORFs prédites ont été utilisées pour les réactions PCR au départ d'une librairie d'ADNc comprenant les séquences de *C. elegans* (cfr. matériel et méthodes pour plus de détails).

Un premier tour de PCR a permis d'amplifier avec succès 642 ORFs. Un second tour d'amplification PCR a été tenté sur les 160 ORFs restantes. En final, il a ainsi été possible de générer des produits PCR pour 764 des 802 ORFs, soit pour 95% du nombre total d'ORFs. Les produits PCR ont été collectés et rassemblés sur 8 plaques 96 puits (WFPCR 020 à WFPCR 027).

Les amplicons PCR sont insérés individuellement dans le vecteur Gateway™ pDONr233 et les produits obtenus sont transformés dans des cellules bactériennes *E. coli* et placées sur

des plaques de milieu bactérien LB (Figure 16). Huit clones bactériens par ORF sont inoculés dans du milieu LB liquide pour subir, après une nuit de croissance à 37°C, une minipréparation d'ADN (cfr. matériel et méthodes). La taille approximative des inserts est vérifiée par PCR et les minipreps d'une même ORF qui présentent un insert de taille attendue sont "poolées".



**Figure 16.** Exemple de plaques de milieu LB (+ ampicilline) sur lesquelles ont été "spottées" les bactéries transformées avec les produits de ligation. Pour chacun des 96 produits PCR d'une plaque, environ 100  $\mu$ L de milieu de transformation a été dispensé. Les figures a et b nous montrent deux boîtes de milieu LB, WFPCR 021 (a) et WFPCR 022 (b), avec des colonies bactériennes après une croissance "overnight". Les colonies bactériennes qui sont capables de croître possèdent donc le plasmide pDONr233 portant, en plus de l'insert ORF, un marqueur de résistance à l'ampicilline.

Sur les 802 ORFs de départ, il a été possible de générer des produits PCR "full length" pour 764 ORFs. 746 d'entre elles (soit 97,6 %) ont été clonées avec succès dans le vecteur pDONr233. Chacun des clones correspondant à une ORF est collecté individuellement dans les plaques 96 puits WFMP 050 à WFMP 057 constituant l'ORFéome de référence. L'étape suivante est de sélectionner pour chacune des ORFs, les clones dits "wild-type", c.-à-d. dont la séquence est identique à celle du génome de référence de *C. elegans*.

## 2. Vérification de la séquence des ORFs de l'ORFéome

Les fragments des 802 protéines essentielles à l'embryogenèse seront générés à partir d'ORFs identiques à celles retrouvées dans le génome de *C. elegans*. Pour les obtenir, les ORFs de notre ORFéome vont être transformées en bactérie et les clones cellulaires présentant l'ORF "wild-type" seront collectés pour former un nouvel ORFéome de clones "wild-type".

### 2.1. Transformation bactérienne<sup>2</sup>

Des bactéries compétentes sont transformées avec les produits des 746 minipreps et "spottées" sur des plaques de milieu LB (cfr. légende de la Figure 17). Pour chacune des ORFs, 4 colonies individuelles (a, b, c et d) sont sélectionnées et stockées dans du glycérol à -80°C avec comme attente qu'au moins une d'entre elles contienne l'ORF "wild-type". Pour limiter dans un premier temps le nombre d'échantillons à séquencer, seules les deux premières colonies sélectionnées (a et b) pour chaque ORF sont amplifiées par PCR et séquencées. La PCR est faite directement sur les bactéries. La taille approximative des inserts est contrôlée sur gel avant le séquençage (Figure 18). Comme les ORFs ont été rangées par ordre de grandeur croissante dans les huit plaques, il est aisé de détecter les pistes sans produit d'amplification ou avec un amplicon de taille anormale (Table 1). Le nombre d'amplicons absents ou anormaux pour la plaque WFMP 057 est d'environ 50% (Figure 18c). Une nouvelle réaction PCR a été menée sur les mêmes clones bactériens a et b. La PCR dans ce cas-ci ne se fait pas directement sur les bactéries, mais après une étape préalable de minipréparation. Ceci a permis d'élever à 70% la proportion d'ORFs ayant un amplicon taille attendue (Figure 18d).

Nom de la plaque 96 puits	Taille moyenne des ORFs (en pb)
WFMP 050	318
WFMP 051	528
WFMP 052	794
WFMP 053	1091
WFMP 054	1386
WFMP 055	1745
WFMP 056	2287
WFMP 057	3349

**Table 1.** Taille moyenne des ORFs par plaque de minipreps. Les 746 minipreps ont été rangées dans les plaques WFMP 050 à 057 selon la taille croissante de leur insert.

<sup>2</sup> Pour les détails techniques complémentaires, cfr. matériel et méthodes

## 2.2. Résultats du séquençage

Le séquençage utilise les amorces M13-G "forward" et "reverse" qui s'hybrident à environ 100 paires de base en amont et en aval de l'insert. Pour chacune des ORFs, deux clones (a et b) ont été séquencés, comme déjà mentionné. On convient de qualifier un clone de "wild-type" si la séquence correspond identiquement à celle du génome (site [www.wormbase.org](http://www.wormbase.org)). Un alignement de séquences est donc réalisé entre la séquence de référence et celle de chacun des clones d'ORFs obtenus en 2.1.

Sur les 746 minipreps transformées en bactéries, on a obtenu des séquences de type "wild-type" pour 527 ORFs (Table 2).

Catégorie	Désignation	Nombre
1.	ORF "wild-type"	527
2.	ORF non "wild-type"	123
3.	ORF à refaire	60
4.	ORF à omettre	36
	<b>TOTAL</b>	<b>746</b>

**Table 2.** Classification des ORFs après le séquençage des deux clones a et b pour chacune des ORF. Le nombre d'ORFs par catégorie sur les 746 minipreps de départ est indiqué (Détails dans le texte).

Les ORFs appelées non "wild-type" (catégorie 2) peuvent l'être pour différentes raisons. Dans environ 10% des cas, la séquence ne correspondait aucunement à celle de l'ORF attendue. C'est généralement le cas des ORFs pour lesquelles l'amplicon PCR était absent sur les gels d'électrophorèse ou ne possédait pas la taille attendue (cfr. Figure 18). Dans environ 60% des cas, on observe une erreur de séquence pour les clones a et b au niveau des amorces d'amplification PCR. Ce qui pourrait s'expliquer par le fait que ces amplicons PCR aient été initiés à partir d'amorces mutées (les amorces sont des produits de synthèse pouvant receler des changements ponctuels).

Pour les 60 ORFs de la catégorie 3, les profils de séquençage donnaient lieu à une lecture ambiguë et ne permettant pas de déceler si on était en présence de la séquence "wild-type" ou non. Le cas typique est celui où pour une même position nucléotidique dans la séquence on observe la présence de deux pics de taille similaire dans les profils de séquençage.

Pour diminuer le nombre de candidats à re-séquencer durant un second round, on convient de considérer que dans les cas où on a affaire à une famille de gènes répétés ou dans le cas de gènes à plusieurs variants d'épissage alternatif, un seul clone "wild-type" est suffisant. C'est la situation de la catégorie 4. 36 ORFs n'ont donc pas été re-testées. L'exemple typique est celui des gènes codant les histones (7 ORFs et 4 ORFs éliminées codent l'histone H2B et H2A respectivement).

Après ce premier séquençage, 527 sur 746 ORFs sont retenues comme "wild-type" (soit 70,6%). Un second round de séquençage est organisé. Les ORFs à re-séquencer correspondent aux 183 ORFs des catégories 2 et 3. Pour ce qui est des 123 ORFs pour lesquelles les clones a et b n'étaient pas "wild-type" (catégorie 2), les clones c et d sont utilisés et permettent pour 34 des 123 ORFs d'identifier des clones "wild-type". Pour les 60 ORFs de la catégorie 3, où on a conservé les mêmes clones a et b, les résultats révèlent que pour 41 des 60 ORFs la séquence au terme du second round est de type sauvage.

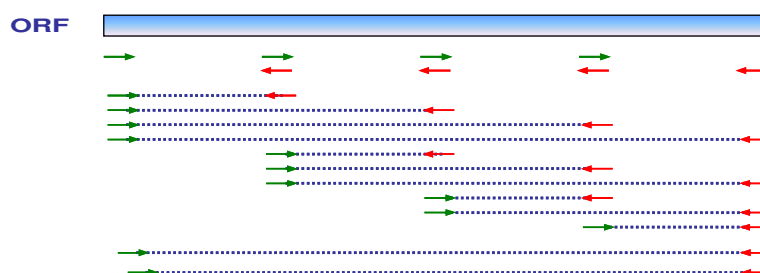
Au total, après le second cycle de séquençage, 602 ORFs "wild-type" sur les 710 ORFs ont été obtenues (84,7%).

### 3. Génération du "fragmentome"<sup>3</sup>

La génération d'une banque de 602 clones "wild-type" de gènes essentiels à l'embryogenèse de *C. elegans* est une ressource précieuse notamment pour la suite de ce travail. En effet, dans le cadre de notre projet, elle sera principalement utilisée pour générer les fragments des différentes ORFs afin de tester les domaines impliqués dans la reconnaissance 2H des protéines de l'embryogenèse de *C. elegans*. Notons que dans un premier temps le projet du "fragmentome" a été initié directement à partir, d'une part, des 764 produits PCR "full length" sans vérification préalable de leur caractère "wild-type" et, d'autre part, de la banque d'ADNc pour la fraction de gènes n'ayant pas de produit d'amplification "full length" (soit 802-764 = 38 ORFs).

#### 3.1. Position et longueur des amorces

Pour générer les fragments d'une ORF, un set d'amorces à intervalles réguliers est conçu. Deux variables sont modulables et ajustées pour obtenir les pièces du "fragmentome" : la taille des fragments et l'intervalle entre les amorces d'un même type ("forward" ou "reverse") (Figure 19). La position des amorces doit évidemment être définie de manière à respecter la phase de lecture de l'ORF.



**Figure 19.** Position des amorces PCR pour générer les fragments d'une ORF. Les flèches vers la droite et vers la gauche correspondent respectivement aux amorces "forward" et "reverse". La ligne en pointillée montre la partie de l'ORF qui sera amplifiée par réaction PCR avec la KOD polymérase (cfr matériel et méthodes). Cette figure ne montre que quelques exemples de fragments générés; un nombre moyen de 40 fragments sera généré par ORF. Dans le bas de l'image sont représentés deux fragments de taille proche du "full length". Ces fragments commencent à la 7<sup>ème</sup> et 13<sup>ème</sup> paires de bases du début de l'ORF, soit respectivement au 3<sup>ème</sup> et 5<sup>ème</sup> codon de l'ORF. Leur ajout au reste des fragments est justifié car il semble, comme mentionné dans l'introduction, que retirer quelques acides aminés d'une protéine est parfois suffisant pour rendre fonctionnelle une fusion protéine/domaine de transactivation qui ne l'était pas avec la protéine "full length".

<sup>3</sup> Dans la suite de ce travail, nous désignerons du terme de "fragmentome" l'ensemble des fragments obtenus à partir des différentes protéines d'un sous-réseau (embryogenèse) destinés à être testés en 2H contre l'une de ces protéines.

Les fragments sont conçus de manière systématique et sans a priori relatif à la prédiction de domaines de structure. La distance des amorces et donc la taille des fragments sont fixées en fonction de la taille totale des ORFs concernées (Table 3). L'idée est de réaliser un nombre de fragments total de l'ordre de 40 quelle que soit la taille totale de l'ORF. Ainsi, d'une part, la taille des fragments est compatible avec la taille de ce qui peut être un "domaine" protéique et, d'autre part, le nombre de fragments reste raisonnable.

Catégorie de taille ORF	Distance entre amorces	Taille des fragments
0-500 pb	75 pb	75, 150, 325, 400 et 475
500 – 1000 pb	100 pb	100, 200, 300, 400, 500, 600, 700 et 800
1000 - 2000 pb	150 pb	150, 300, 450, 600, 750, 1000 et 1500
>2000 pb	200 pb	200, 400, 600, 800, 1400 et 2000

**Table 3.** Distance entre amorces et taille des différents fragments générés par PCR entre paires d'amorces selon la catégorie de taille à laquelle l'ORF appartient.

Le calcul montre que pour les 802 ORFs à fragmenter, il faudra générer par PCR 31804 fragments, soit 40 fragments en moyenne par ORF.

### 3.2. Génération du "fragmentome" pour 44 ORFs

Dans la liste des 802 ORFs essentielles aux premières étapes de l'embryogenèse de *C. elegans*, 44 ORFs de taille moyenne ont été choisies au hasard pour être "fragmentées" (Table 4). Les produits PCR "full length" ont servi de matrice pour les réactions PCR. Pour ces 44 ORFs, il faut générer 1783 fragments et donc réaliser 1783 réactions PCR, réparties sur cinq plaques 384 puits nommées WFPCR 101 à 105. Il a été possible d'obtenir des produits d'amplification pour environ 95% des fragments (Figure 20). Une fois "poolés", les fragments constituent une première "mini-librairie" de fragments que l'on va surnommer "fragmentome 44".

ORF	Nb. Frag	ORF	Nb. Frag	ORF	Nb. Frag
AH6.5	42	C37C3.2a	36	C47E12.1	42
B0334.11a	35	C37C3.2c	36	C52E4.4	36
B0336.6	42	C37H5.6a	41	C53A5.3	42
C30C11.2	48	C37H5.6b	36	C53H9.2a	48
C30F8.2	23	C37H5.8	55	C53H9.2b	29
C31H1.8	54	C38C10.4	49	C54C6.2	35
C33D12.3	49	C39B5.2	35	C55C3.5	49
C33H5.4b	54	C39E9.13	28	D1014.3	29
C34E10.2	28	C39E9.14a	36	DY3.2	54
C34E10.6	49	C40H5.6	29	E01B7.1	36
C36A4.4	42	C41G7.2	54	E02H9.3	35
C36B1.3	35	C41G7.3	54	E03H4.8	29
C36B1.5	41	C43E11.10	49	F01F1.12a	29
C36E8.5	41	C47B2.3	35	F01F1.8a	49
C37A2.4a	49			F01F1.8b	36

**Table 4.** Liste de 44 ORFs, de 1000 pb en moyenne, choisies pour être "fragmentées" par PCR. Le nombre de fragments générés pour chacune de ces ORFs est noté dans la deuxième colonne.



#### **4. Clonage et mise en "pool" des fragments**<sup>4</sup>

On a vu que pour les 802 ORFs environ 30 000 fragments différents seront à tester en 2H contre une protéine "full length". Différents "pools" d'environ 1000 fragments différents seront constitués puis unifiés au moment du test 2H en une librairie de 30000 fragments. On réalise que, compte tenu de la complexité élevée du système d'interactions potentielles d'un tel "pool" comprenant 30000 fragments, il est possible que certaines paires d'interactions réelles échappent à la détection et qu'il faille pour les découvrir réduire, le cas échéant, cette complexité par exemple en utilisant des "sous-pools" de 1000 fragments.

Les fragments sont insérés dans un vecteur double hybride AD-Y par une technique de restriction/ligation. Les produits de ligation sont utilisés pour transformer des bactéries électro-compétentes. Après une croissance "overnight" à 37°C, les colonies bactériennes sont récupérées et subissent une maxipréparation d'ADN. Deux pools d'environ 1000 fragments sont ainsi générés. Un test de diversité de la librairie de fragments a été réalisé sur une des plaques LB de transformation. Il a pour but de s'assurer que certains fragments ne sont pas surreprésentés par rapport à d'autres. 96 colonies ont été récoltées et l'insert contenu dans le vecteur AD-Y a été séquencé. Les fragments séquencés sont effectivement variés, puisque 76,4% d'entre eux ne sont retrouvés qu'en un seul exemplaire ; 20,7% en deux copies et 2,9% en triples copies.

La banque de fragments complète (31804 fragments) n'étant pas encore disponible, un test à échelle pilote a été entrepris avec le "fragmentome 44" avec pour but d'identifier les domaines d'interaction d'interactants (voir ci-dessous) dont on savait qu'à l'état "full length" ils interagissaient en 2H avec certaines des 44 ORFs. Pour mimer au mieux les conditions expérimentales du test de détermination de domaines qui sera effectué avec la banque de 31804 fragments, le "fragmentome 44" (contenant 1783 fragments) est dilué 10x avec une banque de vecteurs d'ADNc-AD. L'idée est ici de déterminer si la sensibilité de la technique est suffisante pour déceler spécifiquement une interaction en conditions grandeur nature du test à haut débit.

---

<sup>4</sup> Pour les détails techniques complémentaires, cfr. matériel et méthodes

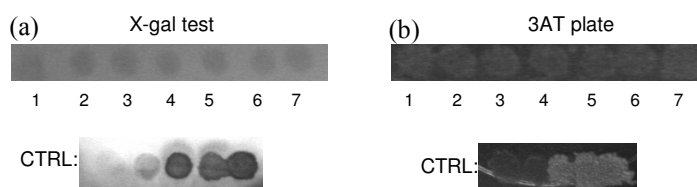
## 5. Utilisation du "fragmentome 44" pour la détection de domaines d'interaction

Dans le but d'éprouver l'hypothèse qu'il est possible de détecter les fragments responsables des interactions paires de 7 protéines appâts "full length" avec 7 autres "full length" en proie, le fragmentome entier, correspondant à 44 protéines dont les 7 détectées en proie, va être testé en 2H contre les 7 protéines appât de départ. (Table 5).

# paire protéique	Protéine appât "full length"	Protéine proie "full length"
1	B0207.4	C30C11.2
2	C02F5.9	C30F8.2
3	C26C6.2	C38C10.4
4	C26F1.4	C53A5.3
5	C28H8.12	C34E10.6
6	C54G10.2	C39E9.13
7	F31E3.3	C39E9.13

**Table 5.** Liste des interactions connues (Worm Interactome 5, Vidal lab). (En réalité, l'interaction #1 ne se retrouve pas dans la version publiée du WI5 mais a été ajoutée par la suite)

En pratique, des levures sont transformées individuellement avec un des sept plasmides double-hybride "appât" où l'ORF de chacune des protéines "full length" est mise en phase avec le domaine de liaison à l'ADN de Gal4. On a vérifié au préalable qu'aucun problème d'auto-activation ne se présentait puisqu'aucune protéine de fusion n'est capable d'activer la transcription des gènes rapporteurs (Figure 21).



**Figure 21.** Tests d'auto-activation avec les rapporteurs lacZ (a) et HIS3 (b) réalisés sur les sept colonies de levure transformées avec les sept constructions DB-ORF différentes (1 à 7). Les contrôles 2H se situent au bas de chaque image (détails dans le texte et matériel et méthodes)

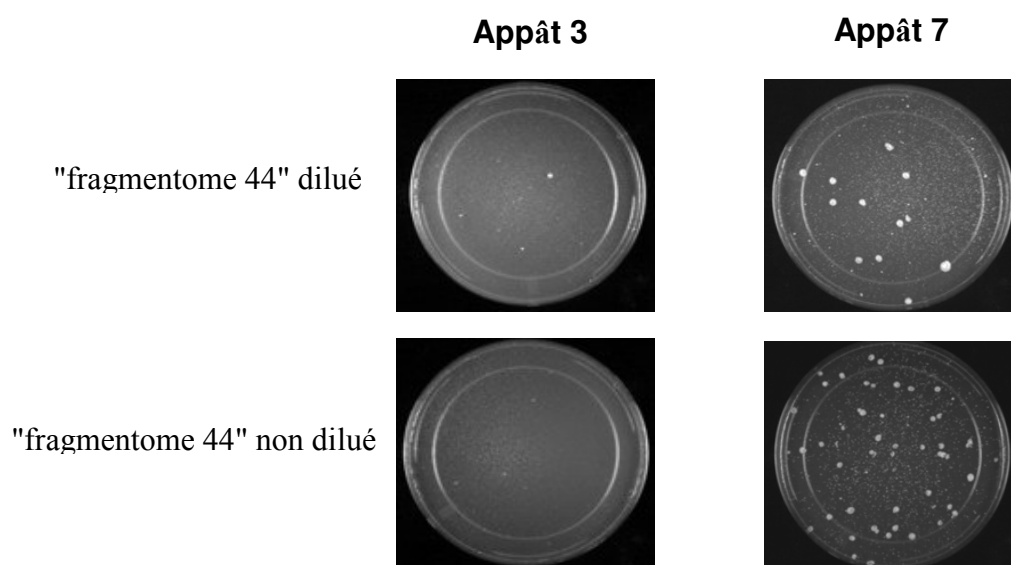
Les 7 colonies de levures sont transformées avec le "fragmentome 44" (Figure 22a). La co-transformation de chacune des constructions DB-ORF est réalisée en double. Une première fois avec la banque de fragments normale et une seconde fois avec la même banque mais diluées 10x dans une librairie d'ADNc-AD (les séquences d'ADNc sont insérées dans un vecteur où elles sont mises en phase avec le domaine d'activation de Gal4) comme indiqué plus haut. Ces deux expériences ("normal" et "dilué") correspondent à des situations où les fragments appartenant aux protéines "full length" de départ sont noyées une première fois avec les fragments des 37 autres protéines et ensuite encore diluées 10x avec des séquences d'ADNc quelconques. Ces deux expériences ont pour but de tenter de simuler la situation telle qu'elle se présentera lorsque le test sera fait en grandeur réelle c.-à-d. lorsque les interactions entre les fragments en proie et les protéines "full length" en appât seront à décélérer parmi une masse de peptides sans partenaires spécifiques. Une fois les co-transformations achevées, les

levures sont étalées sur des plaques 3AT (cfr. matériel et méthodes). En principe, seules les levures montrant une interaction entre un fragment et la protéine "full length" peuvent croître sur ce type de milieu (Figure 22b).

Les résultats montrent que sur les sept transformations effectuées, seules les plaques 3AT où la protéine F31E3.3 a été utilisée comme appât présentent des colonies exploitables. Pour ce qui est des six autres expériences, peu ou pas de colonies de grande taille sont observées, que ce soit avec le "fragmentome 44" dilué ou non (Figure 23). Nous reviendrons sur cette observation plus loin. Dans la suite de ce travail, seule l'interaction 7 sera donc à considérer. Elle implique les partenaires suivants:

- F31E3.3, aussi appelée **rfc-4** : protéine "full length" dans le vecteur BD-X
- C39D9.13, aussi appelée **rfc-3** : protéine dont les fragments sont présents dans le "fragmentome 44".

Une comparaison des plaques 3AT obtenues pour cette interaction nous permet d'observer que dans la condition où le "fragmentome 44" est dilué, le nombre de colonies présentant une interaction est, comme attendu en vertu de la dilution et donc de la diminution de la probabilité qu'un des fragments de rfc-3 soit transformé, plus faible que dans le cas où la banque de fragments ne subit aucune dilution.



**Figure 23.** Comparaison des résultats du test 2H sur milieu 3AT pour les transformants avec le partenaire appât 3 ou 7 et, en proie, le "fragmentome 44" dilué ou non (détails dans le texte). Les photos ont été réalisées 5 jours après la co-transformation. Les plaques des interactions 1, 2, 4, 5 et 6 (non montrées) relatives aux autres paires d'interacteurs "full length" retenus au départ (Table 5) présentent un aspect similaire aux plaques de l'interaction 3.

Les colonies de levures positives sont répliquées à plusieurs reprises sur milieu 3AT comme défini dans le protocole (cfr matériel et méthodes). Un phénotypage des colonies via les tests  $\beta$ -gal et 3AT est réalisé. La majorité des colonies sont positives (Figure 24). Les colonies sont récupérées, lysées et utilisées dans une réaction PCR ayant comme objectif

d'amplifier le fragment compris dans le vecteur AD-Y. La taille des produits PCR est contrôlée comme précédemment avant le séquençage. Pour chaque fragment séquencé, l'identité de l'ORF qui a été utilisée pour le générer est déterminée. On constate que (Table 6) l'utilisation de rfc-4 en "full length" permet de pêcher de manière hautement spécifique (256/261) les fragments de rfc-3 que ce soit avec le "fragmentome 44" dilué ou non.

(a) "fragmentome 44" non dilué

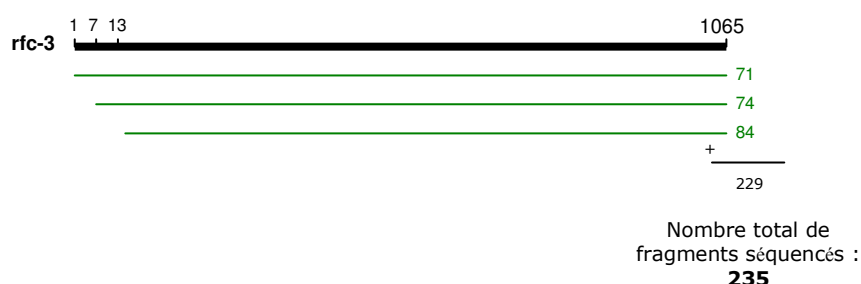
Nom des ORFs	
B0334.11a	1
C31H1.8	1
C34E10.6	1
C39E9.13 (rfc-3)	235
F01F1.8a	1
F01F1.8b	1
<b>Total</b>	<b>240</b>

(b) "fragmentome 44" dilué

Nom des ORFs	
C39E9.13 (rfc-3)	21
<b>Total</b>	<b>21</b>

**Table 6.** Identité des ORFs dont des fragments interagissent en 2H comme "proie" avec la protéine rfc-4 "full length" en "appât" dans la condition où on a utilisé la librairie de fragments diluée (b) ou non (a). La colonne de gauche indique le nombre de fragments retrouvés pour chaque ORF. Seuls les fragments où le séquençage a permis l'identification de la même ORF avec les amorces "forward" et "reverse" sont retenus et comptabilisés.

Le séquençage de fragments donnant une réponse positive en test 2H devait permettre de définir la fraction protéique de rfc-3 indispensable à l'interaction avec rfc-4. Sur les 235 fragments séquencés, trois fragments différents sont majoritairement retrouvés : le fragment allant de la base 1 à 1065, retrouvé 71 fois ; le fragment allant de la base 7 à 1065, retrouvé 74 fois et le fragment allant de la base 13 à 1065, retrouvé 84 fois (Figure 25). L'ORF de rfc-3 ayant une taille de 1065 paires de bases, on voit que les fragments retrouvés correspondent à la protéine quasi complète. Les autres fragments identifiés sont représentés en un ou deux exemplaires et sont des erreurs de séquençage. L'ensemble de la protéine rfc-3 semble donc être indispensable à l'interaction avec rfc-4. Nous reviendrons plus loin sur ce point de discussion.



**Figure 25.** Représentation des fragments de rfc-3 (expérience sans dilution) retrouvés dans le test 2H d'interaction où rfc-4 "full length" est utilisée comme appât. Seuls les fragments où le séquençage a permis l'identification de la même ORF avec les amorces "forward" et "reverse" sont retenus et comptabilisés.

Pour ce qui est du test utilisant la librairie diluée, sur les 21 fragments identifiés par les amorces "forward" et "reverse", le fragment allant de la base 7 à 1065 est retrouvé à 9 reprises et le fragment allant de 13 à 1065 est identifié 4 fois. Les 8 autres fragments identifiés ne le sont qu'une seule fois et se sont avérés être des erreurs de séquençage.

\*  
\*      \*

En résumé, des résultats d'interaction sont obtenus avec le "fragmentome" pour la seule protéine rfc-4 (cas #7), alors que les six autres candidats, pour qui l'interaction 2H en "full length" a aussi déjà été mise en évidence auparavant, ne répondent pas. Ceci paraît paradoxal puisque dans le "fragmentome 44" il y a aussi, par définition, le clone de l'ORF complète ainsi que tous les fragments décroissants par pas d'environ 66 acides aminés.

Remarquons que les interactions entre "fragmentome" et rfc-4 permettent de repêcher des fragments appartenant à des protéines qui étaient reconnues comme partenaires réels 2H "full length" de cette protéine. Ceci est évidemment à porter au crédit de l'hypothèse que la technique "2H fragmentome" est à même de détecter de manière spécifique des fragments ou domaines d'interaction dans un crible comprenant un mélange de grande complexité. Le criblage spécifique en condition de complexité est évidemment la caractéristique exigible pour que la technique soit applicable systématiquement à haut débit.

Notons enfin que rfc-3 n'est pêché par la protéine #7 (rfc-4) qu'à l'état complet ou quasi complet (Figure 25). Les raisons qui pourraient expliquer pourquoi les interactions 2H entre les partenaires "full length" dans les 6 expériences négatives ne sont pas retrouvées et que dans cette 7<sup>ème</sup> expérience "positive" seuls les ORFs quasi complètes interagissent ne sont pas claires et méritent discussion (cfr discussion).

# 1. Discussion

## 1.1 Le nombre d'interactions du test 2H avec le "fragmentome" est faible

Parmi les sept protéines "full length" testées contre le "fragmentome 44", seule la protéine rfc-4 a donné lieu à une interaction. Le fragment proie pêché spécifiquement correspond à l'ORF "full length" de rfc-3. Différentes hypothèses relatives aux constructions (vecteurs-inserts), à la nature du "fragmentome" ou aux protéines appâts elles mêmes ont été émises.

Concernant le vecteur appât (DB-X) utilisé lors des cribles 2H, il est fourni par la firme Invitrogen et ne présente, a priori, aucun défaut. Le vecteur AD-Y où les fragments ont été clonés est un vecteur d'Invitrogen où des sites de restriction AscI et NotI ont été ajoutés dans le site multiple de clonage. Pour des raisons de coût, le clonage des fragments a été réalisé en utilisant une stratégie de restriction/ligation plutôt que le système de recombinaison Gateway™ (longueur des amorces et prix des enzymes plus faibles, cfr matériel et méthodes). Ces deux vecteurs ont fait l'objet de tests contrôle 2H (avant et après le crible 2H) qui prouvent que les vecteurs sont parfaitement compatibles avec des tests 2H.

Le choix des protéines testées en 2H (Table 7) comme proies et appâts a été fait au hasard parmi la vaste liste des 5534 interactions contenues dans le Worm Interactome 5<sup>5</sup> (Li et al. 2004). Cette base de données regroupe différents types d'interactions :

- les interactions identifiées lors d'un test 2H à haut débit à l'échelle du génome entier ("HT-Y2H")
- les interactions identifiées dans la littérature ("litterature")
- les interactions identifiées lors de tests 2H à l'échelle d'un processus biologique individuel ("scaffold")
- les interactions prédites par informatique ("interlog")

# paire protéique	Protéine appât "full length"	Protéine proie "full length"	Type d'interaction
1	B0207.4	C30C11.2	Scaffold
2	C02F5.9	C30F8.2	Scaffold
3	C26C6.2	C38C10.4	Littérature
4	C26F1.4	C53A5.3	Scaffold
5	C28H8.12	C34E10.6	HT-Y2H
6	C54G10.2	C39E9.13	Scaffold + Interlog
7	F31E3.3	C39E9.13	HT-Y2H + Interlog

**Table 7.** Liste des interactions connues (Worm Interactome 5, Li et al. 2004).

Puisque le "fragmentome 44" contient, en plus des fragments, les ORFs "full length", les interactions mises en évidence avec des protéines "full length" dans d'autres tests 2H devraient dès lors être retrouvées dans notre crible.

Une des hypothèses expliquant cette contradiction est que le "fragmentome 44" est peut-être anormalement enrichi en fragments par rapport aux ORFs "full length". Notons que

<sup>5</sup> En réalité, l'interaction #1 ne se retrouve pas dans la version publiée du WI5 mais a été ajoutée par la suite.

l'analyse de l'insert de 96 colonies bactériennes transformantes prises au hasard (cfr point 4 des résultats) ne permet pas de vérifier cette hypothèse.

On peut imaginer que pour les 6 interactions négatives, les protéines de fusion DB-ORF ne soient pas fonctionnelles pour l'interaction. Pour ce qui concerne ces problèmes de clonage et de défaut de phase de l'insert, des contrôles, impliquant le séquençage des quelques clones pris au hasard, ont été effectués et n'ont révélé aucun problème. La croissance des levures sur un milieu sélectionnant la présence du vecteur DB-ORF ne signifie, en effet, pas spécialement que la protéine de fusion soit capable d'interagir en 2H avec la force suffisante pour être détectable, et ce en raison de mutation ou de conformations défavorables résultant de la fusion elle-même.

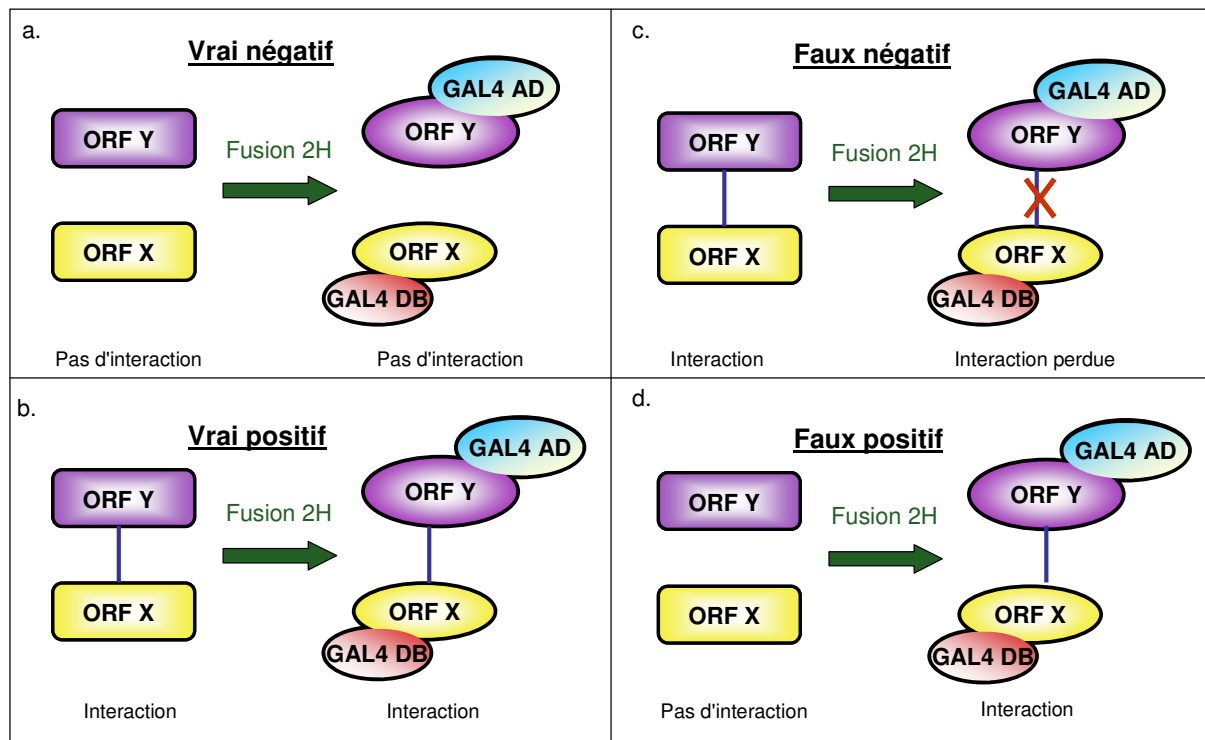
Il est difficile d'estimer la force de liaison entre deux protéines impliquées dans une interaction; néanmoins l'interaction # 7 a déjà été étudiée par d'autres qui l'ont qualifiée d'interaction forte sur base de tests 2H et d'analyses de co-expression (surexpression des deux protéines dans *E. coli* et observation sur gel de la présence d'un complexe soluble stable, cfr Li et Burgers 1994). Un point qui peut argumenter en faveur de la réalité ou non de celle-ci est la prise en considération de la "valeur" de l'interaction prédite. En effet, il faut savoir que pour déterminer si les interactions définies dans le Worm Interactome 5 (WI5) étaient pertinentes du point de vue biologique, elles ont été confrontées aux données des caractérisations phénotypiques (par ARNi) et des profils d'expression (obtenues par puces à ADN dans de nombreuses conditions expérimentales) puis classées en 3 classes de confiance : la classe 1 regroupant les interactions définies par le WI5, les données phénotypiques et les données d'expression ; la classe 2 incluant les interactions du WI5 et les données d'expression uniquement et enfin la classe 3 qui comprend les interactions du WI5 et les données phénotypiques uniquement. Sur les interactions testées dans notre crible, seule l'une d'entre elle appartient à la classe 1 : la # 7. Pour ce qui est des autres, elles ne se retrouvent dans aucune des 3 catégories à l'exception de la # 5 qui est de classe 3. La force et la pertinence biologique de l'interaction # 7 ne fait donc aucun doute. Il n'est pas absurde de considérer que les 6 autres interactions aient été trop faibles pour que l'on puisse distinguer sur les plaques 3AT de transformation (cfr matériel et méthodes) les colonies positives des colonies "background".

Comme dans tout cas de résultats dit "négatif" il est possible d'émettre un grand nombre d'hypothèses explicatives. Toutefois, la stratégie la plus "instructive", à mon sens, est d'éprouver à nouveau cette approche et d'observer si des résultats positifs sont obtenus. Aucune conclusion définitive à cet égard ne peut être retirée du "set" très réduit d'interactions.

## 1.2 Le test n'a été positif que pour le fragment "full length" de rfc-3

L'objectif initial était de déterminer les fragments protéiques de rfc-3 indispensables à une interaction physique avec rfc-4. Seuls les fragments "full length" de rfc-3 ont été retrouvés en 2H. Les interactions avec les fragments sont-elles à considérer comme des faux négatifs ou alors la protéine "full length" est non fragmentable et donc la seule capable d'interagir avec rfc-4 ? Différentes hypothèses ont été avancées pour tenter d'expliquer ce constat.

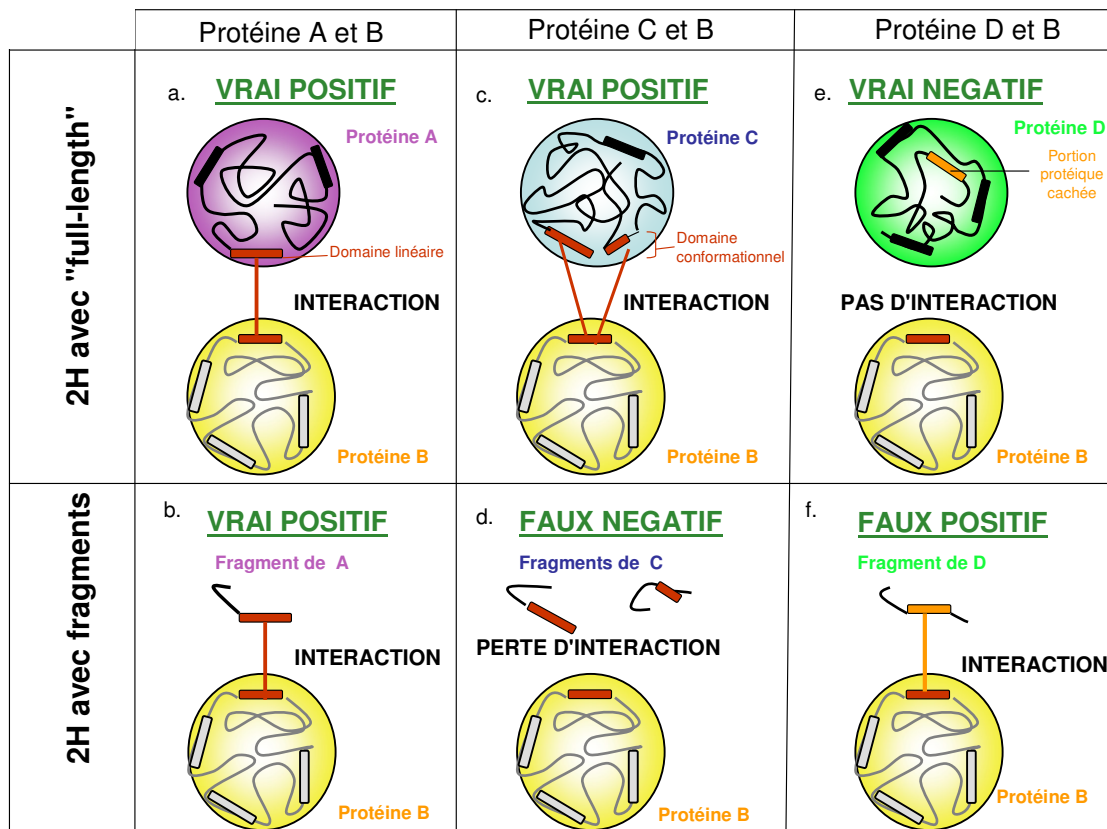
Un paramètre expérimental à ne pas omettre est que dans les tests 2H les deux interactants ne sont pas dans leur conformation originale. En plus du fait que les protéines soient produites dans un autre organisme (dans la levure plutôt que le ver et il est donc possible que des modifications covalentes ou non, notamment d'origine post-traductionnelle, soient observées), elles sont fusionnées aux domaines de liaison à l'ADN ou d'activation de la transcription du facteur Gal4. Cette fusion peut avoir une influence sur la conformation de la protéine ou des fragments, et le cas échéant modifier leur capacité d'interaction (Figure 26). Les fusions "full length" étaient peut-être plus stables que celles des fragments. Cependant, dans certains cas ce sont les constructions impliquant des protéines "full length" qui sont inadaptées aux fusions 2H alors que le retrait de quelques acides aminés est parfois suffisant pour rendre cette fusion fonctionnelle.



**Figure 26.** Représentation schématique des quatre cas qui peuvent se présenter lors d'un test d'interaction 2H. La fusion entre l'ORF étudiée et le domaine AD ou DB induit une modification conformationnelle de la protéine.

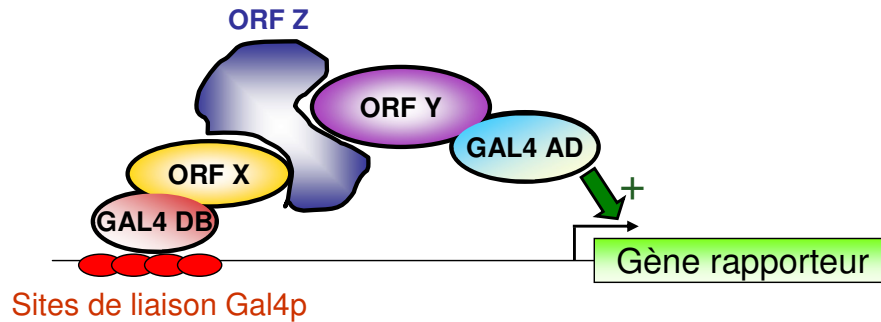


La stratégie du 2H "normal" sur des protéines "full length" génère, comme on vient de le voir, intrinsèquement un certain nombre de faux négatifs et faux positifs (Figure 26). Ceci est évident en considérant tout ce que l'on sait sur la "sensibilité" structurale et les transconformations que manifestent les protéines. L'utilisation de fragments, à la place des protéines complètes, accentue encore plus ce mécanisme en créant des polypeptides qui, en réalité, n'existent pas (Figure 27). Certaines portions protéiques, qui normalement ne sont pas exposées (ou exposées différemment en tout cas) dans la protéine "full length" peuvent constituer des faux positifs en interagissant avec des protéines qui n'auraient pas interagi avec le "full length" ou inversement. (Il n'est possible de détecter aucun domaine conformationnel en utilisant les fragments, seule la détermination des domaines linéaires est accessible avec ce type d'approches.)



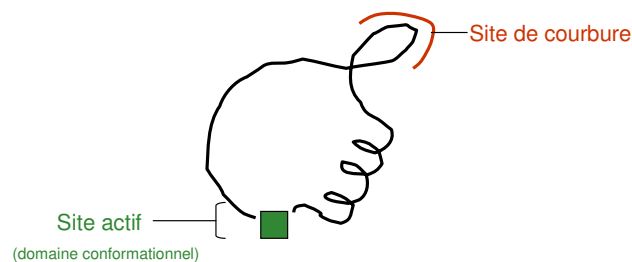
**Figure 27.** Représentation schématique de différents cas d'interaction 2H entre deux protéines "full length" d'une part (a, c et e) et entre une protéine "full length" et des fragments protéiques d'autre part (b, d et f). Le domaine d'interaction conformationnel de la protéine C, qui permet en "full length" l'interaction avec la protéine B (c), est non fonctionnel si on utilise les fragments. Par contre la plupart des domaines linéaires restent fonctionnels même isolés en fragments (a et b). On s'attend à ce que des portions protéiques enfouies dans la structure tertiaire (e) peuvent, grâce au processus de fragmentation, interagir avec des protéines qui ne leur sont normalement pas accessibles et constituer ainsi des faux positifs.

Bien d'autres cas de faux positifs ou faux négatifs peuvent être encore envisagés, citons par exemple le cas où les deux protéines testées interagissent via l'intermédiaire d'une troisième protéine non connue (cas de triple-hybride (Figure 28)). En résumé, il faut retenir que le 2H reste une stratégie permettant l'identification de nouvelles interactions potentielles qui seront considérées réellement pertinentes biologiquement que si et seulement si elles ont été confirmées par d'autres approches.



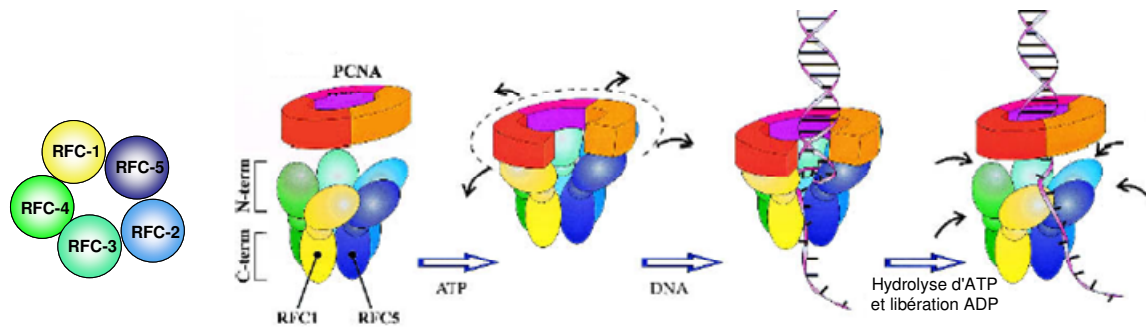
**Figure 28.** Représentation d'une interaction entre trois partenaires qui constituera un faux positif dans un test 2H. Une interaction entre X et Y sera déduite alors qu'en réalité ces deux protéines n'interagissent pas physiquement. Par contre, dans un test triple-hybride, un cas de faux négatif est observé si cette protéine Z (ou protéine homologue) est absente dans la levure.

Le "folding" des fragments est évidemment à considérer. Les protéines étant des macromolécules complexes, une délétion au sein de la séquence primaire peut affecter directement la séquence d'interaction ou affecter indirectement l'interactivité. Par exemple, on sait qu'un "folding" peut dépendre de modifications à grande distance (indirecte) des résidus directement impliqués dans l'interaction. Donc, dans une protéine, les acides aminés importants ne sont pas uniquement ceux qui appartiennent aux domaines protéiques; les acides aminés reliant différents domaines peuvent être essentiels à l'obtention d'une conformation tertiaire ou quaternaire adéquate (cfr par exemple Figure 29). Si plusieurs portions de la protéine présentent ce type de caractéristiques, la protéine est difficilement divisible en fragments à structure stable ayant maintenu l'aptitude d'interagir et dans ce cas seule l'interaction avec la protéine "full length" pourra être retrouvée. La stabilité structurale est fonction de la séquence et du "folding" des différents fragments protéiques. Il est difficile de tester spécifiquement la stabilité d'une conformation dans le cadre d'une expérience à aussi haut débit. Même une mesure physique de la stabilité ne refléterait pas la situation du test 2H qui conduit – on le sait – à des faux positifs et des faux négatifs. (Une analyse de contrôle électrophorétique, par exemple, demanderait que chacun des fragments (+ de 30000 !) soit traduit *in vitro/vivo* et chargé sur des gels à haute résolution, sans qu'on soit assuré d'étudier la conformation native, active du peptide au sein de la protéine "full length".)



**Figure 29.** Représentation schématique d'une protéine. Deux structures sont mises en évidence : le site actif et le site de courbure indispensable au "folding" adéquat de la protéine. Retirer les acides aminés de l'un de ces deux sites implique directement une perte de fonction du site actif.

A côté de ces hypothèses à caractère plutôt "technique", il est possible d'avancer des hypothèses pour envisager une cause "biologique". Les protéines rfc-3 et rfc-4 font partie intégrante d'un complexe essentiel impliqué dans la réplication et la réparation de l'ADN (replication factor C) (Figure 30). On pourrait imaginer que, pour des raisons qui restent à déterminer, ces protéines évoluent de manière à ce que la totalité de leur chaîne peptidique devienne le domaine d'interaction. La protéine pourrait dès lors être considérée comme un "domaine" à part entière dans le complexe, ce qui expliquerait pourquoi le test 2H avec le "fragmentome" n'a identifié que le fragment "full length".



**Figure 30.** (a) Arrangement schématisé des sous-unités dans le complexe RFC. (b) Le facteur de réplication c (RFC) est un complexe protéique hétéropentamérique qui intervient dans la réplication et la réparation de l'ADN. Pendant la réplication de l'ADN, le complexe RFC, en collaboration avec le PCNA (proliferating cell nuclear antigen), forme des protéines accessoires de l'ADN polymérase eucaryotique. RFC lie l'ADN et utilise l'ATP pour amener le PCNA sur l'ADN. PCNA est une protéine homotrimérique qui entoure l'ADN et attache l'ADN polymérase sur l'ADN "template" pour lancer la synthèse. (Venclovas et al. 2002)

En conclusion, il n'a pas été possible de définir avec précision le domaine d'interaction de rfc-3 mais il est plausible que ce cas ne soit pas exemplatif de la situation des autres protéines (telle que TIR-1) du protéome. En effet, des indices comme le fait que seuls les fragments proches du "full length" aient été pêchés nous poussent à croire que rfc-3 n'est pas une protéine "fragmentable". Ceci peut-être dû à des raisons de stabilité (les fragments de la protéine ne sont pas stables) ou encore soit au fait que son domaine d'interaction avec rfc-4 est de type conformationnel.

Pour évaluer comment les protéines se répartissent en "fragmentables" ou non "fragmentables" il faudrait disposer d'expériences du type de celle que nous avons décrite ici.

## 2. Perspectives

La génération des fragments d'ORFs est réalisée de manière systématique et sans a priori grâce à des amorces PCR disposées toutes les  $n$  paires de bases (avec respect de la phase). Dans un certain nombre de cas, il faut s'attendre à ce qu'une amorce puisse venir s'hybrider à l'intérieur d'une structure secondaire protéique (hélice  $\alpha$  ou feuillet  $\beta$ ), générant ainsi des fragments qui risquent de coder des polypeptides de faible stabilité. Alors que pour définir le domaine d'interaction de TIR-1 avec la protéine KGB-1 (cfr introduction), plus de 100 fragments différents de TIR-1 avaient été utilisés, dans notre approche, on utilise une moyenne de seulement 40 fragments par ORF, ce qui implique que la perte éventuelle de fragments suite à l'instabilité de certaines structures secondaires est beaucoup plus pénalisante. Pour ne pas devoir augmenter de manière excessive le nombre total de réactions PCR à faire pour l'ensemble des 802 ORFs, une stratégie basée sur la prédiction de structures secondaires pourrait être appliquée. De manière systématique, les amorces pourraient être positionnées après chaque structure secondaire prédite, mais bien entendu dans cette hypothèse l'approche est amputée, en partie, de son caractère systématique et sans a priori.

Un contrôle de la normalisation du "fragmentome" afin de s'assurer que certains fragments ne sont pas surreprésentés par rapport à d'autres pourrait se révéler être intéressant. Cependant, le fait que l'interaction # 7 soit positive est en soi un indice de qualité de la banque utilisée.

Une autre chose à faire pour tester l'efficacité du système serait de disposer de contrôles positifs, c.-à-d. d'utiliser des protéines pour lesquelles les domaines d'interaction ont déjà pu être déterminés grâce à l'approche du "fragmentome" (par exemple le cas des ORFs de LIN-5 et de GOA-1 (cfr figure 13)).

Il va de soi, mais il y a lieu d'y insister, que les informations apportées par ce type d'analyse 2H devront être, comme toute donnée obtenue par le haut débit, validées par d'autres sources. Le 2H fournit des "candidats" et par là des hypothèses qui devront être validées. Comme perspectives, il pourrait par exemple être possible d'imaginer des tests d'interaction peptide-peptide *in vitro* dans des matrices de grande taille. Les interactions seraient mises en évidence par différentes techniques permettant de détecter la formation d'un complexe entre les deux peptides (gel d'électrophorèse...). Les "protein-chips" ou "peptides-chips" pourraient être exploitable dans ce contexte. On pourrait également tenter de prédire les structures correspondant à un "fragmentome". L'accumulation de données et la comparaison de domaines interagissant, c.-à-d. des structures paires sensées assurer une interaction biologique, et des séquences correspondantes devraient permettre de dégager progressivement les lois régissant ces interactions et permettre ensuite, peut-être, de les prédire sur base d'alignements de séquences de différents "fragmentomes". Enfin, une stratégie basée sur des mutations ponctuelles est envisageable. L'ORF d'une des deux protéines partenaires serait la cible de mutations ponctuelles faites individuellement. Après chaque mutation, l'interaction entre les deux partenaires est testée et quantifiée (qualitativement et/ou quantitativement). Une perte d'interaction signifie que l'acide aminé muté affecte le domaine d'interaction (Figure 31). Mais, on pourrait imaginer des cas où l'interaction est perdue non pas à cause d'une perturbation d'un domaine mais à cause d'une perte de stabilité globale de la protéine. Une expérience de mutation compensatoire, supposée rétablir l'interaction, sur la protéine partenaire B peut nous permettre de confirmer l'identité des domaines en interaction et de contrôler en même temps la stabilité de la protéine A.

### **3. Conclusions**

Ce test n'est qu'une expérience pilote d'un test à effectuer sur l'ensemble des protéines essentielles à l'embryogenèse de *C. elegans*. En d'autres termes, il s'agit de définir la stratégie la plus appropriée pour convertir une approche qui a fait ses preuves en bas débit (TIR-1, cfr introduction) en une approche applicable aux analyses à grande échelle. Une fois correctement configuré, le "HT-fragmentome" (High-Throughput Fragmentome) pourrait être applicable aussi bien aux gènes de l'embryogenèse de *C. elegans* qu'à l'ensemble des gènes de son génome, voire des génomes plus complexes tel que celui de l'humain.

Nos résultats montrent que la technique possède une sensibilité suffisante pour pouvoir retrouver les fragments interagissant avec une protéine "full length" en appât dans une banque complexe c.-à-d. en conditions réelles d'un crible à haut débit. En effet, même avec le "fragmentome 44" dilué 10x (pour simuler la présence de nombreux fragments différents), il a été possible de retrouver spécifiquement les fragments de rfc-3 en utilisant rfc-4 comme appât. Ceci établit un point important pour la faisabilité de l'approche.

D'autres expériences seront nécessaires pour tester la possibilité de définir par fragmentation les domaines spécifiques d'interaction. En effet, il semble que les protéines se répartissent en deux classes : les protéines dont on peut générer des fragments stables et fonctionnels et les protéines qui ne fonctionnent qu'à l'état "full length", à laquelle semble appartenir la protéine rfc-3.

# **1. Restriction/ligation d'ADN**

## **1.1. Enzyme de restriction**

### **1.1.1. Restriction de produits PCR**

Les restrictions sont réalisées après une purification avec le kit QIAquick (Qiagen Catalogue Nr 28104) dans un volume final d'élution de 80 µL.

#### **Réaction de restriction :**

- 75µL de produits PCR purifiés
- 10 µL tampon NE (le type de tampon est variable en fonction de l'enzyme utilisée)
- 4 µL d'enzyme de restriction + 100 µg/mL BSA si nécessaire
- Ajouter de l'eau distillée pour atteindre un volume final de 100 µL.

Incubation à 37°C pendant 3 heures

### **1.1.2. Restriction du vecteur pPC86 modifié**

Le vecteur pPC86 (Figure 32) a été légèrement modifié :

- les sites de restriction AscI et NotI ont été ajoutés dans le site multiple de clonage

#### **Réaction de restriction :**

- 10 µg de plasmide
- 10 µL tampon NE (le type de tampon est variable en fonction de l'enzyme utilisée)
- 4 µL d'enzyme de restriction + 100 µg/mL BSA si nécessaire
- Ajouter de l'eau distillée pour atteindre un volume final de 100 µL.

Incubation à 37°C pendant trois heures. Dans la dernière heure, ajouter 2 µL de phosphatase intestinale de veau (Calf Intestinal Phosphatase 10000 unités/mL) qui permet d'éviter une re-circularisation du vecteur vide.

La Table 8 reprend les différentes enzymes de restriction utilisées lors de ce travail. Toutes les enzymes de restriction proviennent de la firme *New England Biolabs*.

## **1.2. Ligation d'un insert dans un vecteur :**

- Vecteur et insert dans un rapport 1 : 10
- 1 µL du tampon (10X) de réaction de la T4 DNA Ligase
- 1 µL T4 DNA Ligase 2000000 unités/mL
- Ajouter de l'eau distillée pour atteindre un volume final de 10 µL.

Incuber "overnight" à 16°C.

## **1.3. Utilisation des enzymes de restriction/ligation pour cloner les fragments:**

Pour que les fragments soient utilisables en 2H, il faut qu'ils soient insérés dans un vecteur où la séquence codante du fragment est en phase avec le domaine de transactivation de GAL4. Cette insertion dans le vecteur AD-Y est faite par une technique de restriction/ligation. Chacune des amorces "forward" et "reverse" utilisées pour générer les fragments possédaient en 5' une séquence comprenant le site de restriction de AscI et NotI respectivement. Les fragments de chacune des 5 plaques 384 puits sont collectés dans un tube eppendorf. Le contenu de chaque tube eppendorf subit une étape de restriction avec les deux enzymes suivie d'une étape de ligation dans le vecteur AD-Y (pPC86 AscI/NotI) avec la T4 DNA ligase. Les produits de ligation sont utilisés pour transformer dans des bactéries électro-compétentes et étalés sur 5 plaques LB de 15 cm de diamètre (+ampicilline). Les colonies des plaques LB sont raclées et lavées avec du LB liquide puis stockées dans un erlen. Deux erlens sont utilisés, le premier regroupe les colonies des plaques LB 101, 102 et 103 et le second

regroupe celles des plaques 104 et 105. Deux maxipréparations sont réalisées séparément sur les deux cultures générant ainsi deux pools de fragments d'une concentration d'environ 0,3 µg/µL chacun. Ils comprennent chacun environ 1000 fragments différents.

## **2. Production, purification et analyse de produits PCR**

### **2.1. Réaction PCR**

La Polymerase Chain Reaction (PCR) consiste en une amplification exponentielle d'un fragment d'ADN à l'aide d'amorces (petites séquences oligonucléotidiques) et d'une enzyme, l'ADN polymérase. Mise au point en 1985 par Karry Mullis, elle permet de repérer un fragment d'ADN même présent en quantité infime dans un mélange, puis de le multiplier rapidement. La PCR utilise de manière répétitive la propriété des ADN polymérases : la synthèse d'un brin complémentaire d'ADN à partir d'une amorce. La PCR repose sur trois étapes répétées plusieurs fois : la dénaturation, l'hybridation et l'élongation.

#### **2.1.1. Invitrogen Platinum® Taq ADN polymerase High Fidelity**

Réactions PCR pour amplifier les séquences nucléotidiques avant le séquençage. La Table 9 donne les volumes standard de ce type de réaction PCR.

#### **2.1.2. Novagen/Toyobo KOD Hot Start ADN Polymérase**

L'ADN polymérase KOD Hifi, activable via la chaleur, permet une augmentation sensible de la spécificité et une préparation plus simple de la réaction PCR. Cette ADN polymérase constitue en réalité un complexe entre l'ADN polymérase et deux anticorps monoclonaux qui inhibent son activité exonucléase 3'→5' à température ambiante et donc la dégradation des amorces avant la réaction PCR. L'amplification non spécifique est réduite car les événements de "mispriming" (mésamorçage) qui peuvent avoir lieu durant la préparation et l'augmentation initiale de température sont évités. Grâce à ses propriétés de haute fidélité, cette enzyme est particulièrement recommandée pour une amplification sans erreur d'une séquence nucléotidique. La Table 10 donne les volumes standard de ce type de réaction PCR.

Pour générer les ORFs "wild-type" des gènes essentiels à l'embryogenèse de *C. elegans*, la librairie d'ADN complémentaire Proquest (Invitrogen), a été utilisée comme "template". Le programme PCR utilisé comporte 30 cycles et un temps d'élongation défini en fonction de la taille moyenne des ORFs. Une vérification approximative de la taille des amplicons est réalisée en utilisant des gels d'agarose "E-gel" adaptés au travail en haut débit.

#### **2.1.3. Amorces utilisées et programmes PCR**

La Table 11 reprend les différentes amorces utilisées durant ce travail. Pour la génération des produits PCR "full length" des 802 ORFs, la séquence des amorces "forward" et "reverse" des ORFs a été déterminée par analyse informatique. La Figure 33 montre un programme PCR standard tel qu'il a souvent été utilisé. Cependant, le nombre de cycles et le temps d'élongation peuvent varier en fonction des conditions expérimentales.

### **2.2. Purification des produits PCR via le kit QIAquick :**

Le kit de purification PCR QIAquick (Qiagen Catalogue Nr 28104) contient une colonne capable de fixer l'ADN (100 pb à 10 kb). Cette purification est idéale pour retirer les amorces,

les nucléotides, les enzymes, les sels, l'agarose, le bromure d'éthidium et d'autres impuretés des échantillons d'ADN. Le protocole disponible on-line sur le site de Qiagen a été suivi.

### **2.3. Electrophorèse en gel d'agarose**

Méthode de séparation de fragments d'ADN, dans un champ électrique, en fonction de leur taille. La révélation de la présence d'ADN est réalisée grâce au bromure d'éthidium qui rend l'ADN visible aux UV en s'y intercalant.

Gels pré-coulés et adaptés à la réalisation d'électrophorèses à haut débit. La résolution permet de séparer et visualiser des produits PCR entre 100 pb et 10 kb. Le 2% E-gel® 48 Agarose (Invitrogen catalogue G8008-02) est utilisé pour visualiser la présence de produits PCR "full length" des ORFs. Par gel, il est possible de faire migrer 48 échantillons. Par contre, le 1% E-Gel® 96 Agarose (Invitrogen catalogue G7008-01), qui contient 96 pistes de migration, est utilisé pour visualiser la présence de produits PCR des fragments. Par puits, 5 µL de produits PCR est additionnés à 10 µL d'eau distillée. Le marqueur de poids moléculaire utilisé est l'E-Gel® High Range DNA Marker (Invitrogen catalogue 12352-019).

## **3. Préparation d'ADN bactérien**

### **3.1. Minipréparation avec le kit QIAprep Spin Miniprep**

Ce kit est conçu pour isoler jusqu'à 20 µg d'ADN plasmidique (Qiagen Catalog Nr 27104). Le protocole disponible on-line sur le site de Qiagen a été suivi.

### **3.2. Minipréparation avec plaques à 96 puits profonds**

Les bactéries sont inoculées dans 1 mL de milieu LB liquide. Après une nuit d'agitation à 37°C, une minipréparation d'ADN est effectuée à haut débit par le robot Qiagen Bio Robot 8000 suivant le protocole fourni par Qiagen. La taille approximative de l'insert est vérifiée par PCR.

### **3.3. Maxipréparation avec le kit Hispeed Plasmid Midi**

Kit permettant de réaliser une préparation d'ADN à plus large échelle (Qiagen Catalog Nr 12643). Le protocole disponible on-line sur le site de Qiagen a été suivi.

## **4. Milieux de culture utilisés**

### **4.1. Milieu bactérien**

#### **4.1.1. Milieu S.O.C. (Invitrogen Catalog No. 15544-034)**

<b>Composition</b>	2% Tryptone	0.5% extrait de levures
	10 mM NaCl	2.5 mM KCl
	10 mM MgCl <sub>2</sub>	10 mM MgSO <sub>4</sub>
	20 mM glucose	
<b>Application</b>	Le milieu S.O.C. (Salt Optimized Carbon) est idéal pour obtenir lors de la dernière étape d'une transformation bactérienne une efficacité de transformation maximale (Hanahan 1983).	



#### 4.1.2. Milieu LB (Luria-Broth) solide et liquide

La technique la plus simple est de préparer 2L de milieu. Une première flasque est remplie avec 1L d'eau et 40 g d'agar. Une autre flasque contient 1L de milieu LB 2x concentré. Après l'étape d'autoclave, les deux flasques seront réunies pour atteindre un volume de 2L.

Milieu LB 2x concentré (1L):	4% Agar (1L):
<ul style="list-style-type: none"> <li>▪ Placer dans une flasque de 2L:               <ul style="list-style-type: none"> <li>• 20 g Bacto Tryptone (Difco #0123-17-3)</li> <li>• 10 g d'extrait de levures (Difco #0127-17-9)</li> <li>• 10 g NaCl</li> </ul> </li> <li>▪ Ajouter de l'eau pour atteindre le volume de 1L.</li> <li>▪ Autoclaver pendant 45 minutes (cycle liquide)</li> </ul>	<ul style="list-style-type: none"> <li>▪ Placer dans une flasque de 2L :               <ul style="list-style-type: none"> <li>• 40 g Difco Agar (Difco #214010)</li> </ul> </li> <li>▪ Ajouter de l'eau pour atteindre le volume de 1L.</li> <li>▪ Autoclaver pendant 45 minutes (cycle liquide).</li> </ul>

#### **Pour préparer le milieu :**

Après l'autoclave, ajouter les éléments suivants à la flasque milieu LB :

1. L'antibiotique (si souhaité). Ajouter uniquement après que le milieu soit refroidi.
  - 2.0 mL Ampicilline (50 µg/mL)
  - 2.0 mL Kanamycine (50 µg/mL)
  - 2.0 mL Chloramphénicol (50 µg/mL)
2. Pour faire du milieu liquide, ajouter de l'eau stérile pour atteindre le volume de 2L.
3. Pour faire du milieu solide, ajouter le litre d'agar 4% à la flasque de milieu LB. Mélanger avec un barreau magnétique. Couler dans des boîtes de Pétri stériles de 10 cm.

#### 4.2. Milieu pour levures

##### 4.2.1. Milieu non sélectif : milieu YEPD (Yeast Extract Peptone Dextrose)

La technique la plus simple est de préparer 2L de milieu. Une première flasque est remplie avec 1L d'eau et 40 g d'agar. Une autre flasque contient 900 mL de milieu YEPD 2x concentré. Après l'étape d'autoclave, 100 mL de glucose 40% sont ajoutés à une des deux flasques avant qu'elles ne soient réunies pour atteindre un volume final de 2L.

Milieu YEPD 2x concentré (1L):	4% Agar (1L):
<ul style="list-style-type: none"> <li>▪ Placer dans une flasque de 2L:               <ul style="list-style-type: none"> <li>• 20 g d'extrait de levures (Difco #0127-17-9)</li> <li>• 40 g de bactopeptone (Difco #0118-17)</li> <li>• Porter avec de l'eau à un volume de 900 mL</li> </ul> </li> <li>▪ Ajouter de l'eau pour atteindre le volume de 1L.</li> <li>▪ Autoclaver pendant 45 minutes (cycle liquide)</li> <li>▪ Ajouter 100 mL de glucose 40%.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Placer dans une flasque de 2L :               <ul style="list-style-type: none"> <li>• 40 g Difco Agar (Difco #214010)</li> </ul> </li> <li>▪ Ajouter de l'eau pour atteindre le volume de 1L.</li> <li>▪ Autoclaver pendant 45 minutes (cycle liquide).</li> </ul>

#### **Pour préparer le milieu :**

Après l'autoclave, ajouter les éléments suivants à la flasque YEPD 2x et mélanger :

1. Pour faire du milieu liquide : ajouter de l'eau stérile jusqu'à atteindre le volume de 2L
2. Pour faire du milieu solide, ajouter le litre d'agar 4% à la flasque de milieu YEPD. Mélanger avec un barreau magnétique. Couler dans des boîtes de Pétri stériles de 15 cm.

##### 4.2.2. Milieu sélectif :

La technique la plus simple est de préparer 2L de milieu. Une première flasque est remplie avec 1L d'eau et 40 g d'agar. Une autre flasque contient 900 mL de milieu SC 2x concentré. Après l'étape d'autoclave, 100 mL de glucose 40% sont ajoutés à une des deux flasques avant qu'elles ne soient réunies pour atteindre un volume de 2L.

Milieu SC 2x concentré (1L):	4% Agar (1L):
<ul style="list-style-type: none"> <li>▪ Placer dans une flasque de 2L:               <ul style="list-style-type: none"> <li>• 3,4 g de Yeast Nitrogen Base (YNB) (Dfci #0355-15-9)</li> <li>• 10 g de sulfate d'ammonium (Sigma #A2939)</li> <li>• 2,6 g de poudre d'acides aminés (-Trp, -His, -Leu, -Ura).</li> </ul> </li> <li>▪ Ajouter de l'eau pour atteindre le volume de 900 mL</li> <li>▪ Ajuster le pH à 5,9 (utiliser le NaOH)</li> <li>▪ Autoclaver pendant 45 minutes (cycle liquide)</li> <li>▪ Ajouter 100 mL de glucose 40%.</li> </ul>	<ul style="list-style-type: none"> <li>▪ Placer dans une flasque de 2L :               <ul style="list-style-type: none"> <li>• 40 g Difco Agar (Difco #214010)</li> </ul> </li> <li>▪ Ajouter de l'eau pour atteindre le volume de 1L.</li> <li>▪ Autoclaver pendant 45 minutes (cycle liquide).</li> </ul>

### Pour préparer le milieu :

Après l'autoclave, ajouter les éléments suivants à la flasque SC 2x et mélanger :

1. Pour faire du milieu liquide : ajouter de l'eau stérile jusqu'à atteindre le volume de 2L
2. Pour faire du milieu solide, ajouter le litre d'agar 4% à la flasque de milieu SC. Mélanger avec un barreau magnétique. Couler dans des boîtes de Pétri stériles de 15 cm.

En fonction du type de plaque sélective à réaliser, ajouter, avant de verser dans les boîtes de Pétri, 16 mL des solutions suivantes : 100 mM histidine-HCl, 20 mM uracile, 100 mM leucine ou 40 mM tryptophane. Ces solutions peuvent être conservées à température ambiante à l'exception du tryptophane qui doit être stocké dans le noir à 4°C. Des drogues comme le 3-amino-triazole (3,36 g 3AT pour 2L de milieu; concentration finale de 20 mM 3AT) et/ou le cycloheximide (200 µL cycloheximide 10 µg/mL pour 2L de milieu) peuvent être ajoutées.

**Glucose (40%) :** Dissoudre 400 g de glucose dans un volume final de 1L d'eau distillée. Autoclaver la solution pendant 45 minutes.

### **Poudre d'acides aminés**

Pour préparer la poudre:

1. Mélanger 6 g de chacun des acides aminés suivants: alanine, arginine, acide aspartique, cystéine, acide glutamique, glutamine, glycine, isoleucine, lysine, méthionine, phénylalanine, proline, sérine, thréonine, tyrosine et valine.
2. Ajouter 6 g de sulfate d'adénine.

La poudre contient des acides aminés et d'autres facteurs nécessaires à la croissance de levures suivant des conditions définies.

## **5. Transformation**

La transformation est une technique qui permet de faire rentrer de l'ADN exogène (plasmide) dans une cellule.

### **5.1. Transformation bactérienne**

#### **5.1.1. Bactéries thermocompétentes:**

Pour ce type de transformation, les bactéries *E. Coli* One Shot® OmniMAX™ 2 T1 Phage-Resistant Cells ont été utilisées (Invitrogen Cat. No. C8540-03).

**Procédure de transformation :**

1. Placer 100 µL de bactéries thermocompétentes dans un eppendorf sur glace.
2. Une fois les bactéries décongelées, ajouter 1 à 5 µL d'ADN (10 pg à 100 ng).
3. Incuber sur glace pendant 20 minutes
4. Choc thermique à 42°C pendant 45 secondes.
5. Remettre les bactéries sur glace pendant 2 minutes.
6. Ajouter 250 µL de milieu S.O.C. pré-chauffé.
7. Agiter à 225 rpm dans une pièce à 37°C pendant une heure.
8. Etaler les bactéries sur un milieu sélectif approprié.
9. Inverser la plaque et incuber "overnight" à 37°C.

Pour générer les clones "wild-type" ce protocole a été légèrement adapté. Travaillant avec des plaques 96 puits, les chocs thermiques sont réalisés en utilisant une machine PCR (MJ Research PTC 225 DNA engine tetrad PCR machines), 100 µL de milieu bactérien S.O.C. est ajouté à chaque puits, les plaques une fois scellées sont placées à une température de 37°C pendant 60 minutes.

**5.1.2. Bactéries électrocompétentes**

Les bactéries ElectroMAX DH5α-E ont été utilisées (Invitrogen Cat. No. 11319-019).

**Procédure de transformation :**

1. Ajouter 3 µL d'ADN dans un tube de micro-centrifugation.
2. Mettre les cellules ElectroMAX DH5α-E. sur glace
3. Une fois les cellules décongelées, ajouter 50 µL de cellules dans le tube de micro-centrifugation.
4. Placer la mixture ADN/cellules dans une cuvette spécifique et réaliser l'électroporation avec les conditions suivantes : 1,8 kV, 200 Ω, 25 µF (BioRad Gene Pulser® II electroporator).
5. Ajouter dans la cuvette 1 mL de milieu S.O.C. et transférer la solution dans Falcon tube de 15 mL.
6. Agiter pendant une heure à 37°C (225 rpm)
7. Etaler 300 µL de bactéries sur un milieu sélectif approprié (plaque Pétri de 15cm)
8. Inverser la plaque et incuber "overnight" à 37°C.

**5.1.3. Stockage de bactéries sous forme de stock glycérol**

Après une transformation, les colonies de la plaque LB sont utilisées pour inoculer 100 µL de LB liquide (+ antibiotique) contenu dans une plaque Costar de 96 puits (dont les puits sont en forme de U, Corning Catalog No. 3799). Après une croissance "overnight" à 37°C, 100µL de glycérol 40% est ajouté pour permettre le stockage des bactéries à -80°C.

**5.2. Transformation de levures**

Protocole pour transformer les levures d'une plaque 96 puits.

- Faire croître les levures "overnight" à 30°C, sous agitation (250 rpm) dans 100 mL de milieu YEPD liquide. Le lendemain matin vérifier l'absorbance à 600 nm ( $A^{600}$ ) doit être entre 0,1 et 0,5.
- Préparer :
  - l'ADN de sperme de saumon (Sigma Catalog No. D-9156, 10µg/µL). Mettre le tube d'ADN de sperme de saumon au bain-marie à 100°C pendant 10 minutes. Puis, mettre sur glace.
  - la solution de TE/LiAc
  - la solution de TE/LiAc/PEG
- Verser 50 mL de la culture de levures dans un Falcon tube de 50 mL (Falcon Cat. No. 352098).
- Centrifuger 3 minutes à 1800 rpm puis resuspendre avec 50 mL d'eau stérile.
- Centrifuger 3 minutes à 1800 rpm puis resuspendre avec 10 mL de TE/LiAc.
- Centrifuger 3 minutes à 1800 rpm puis resuspendre dans un volume total de 0,2 mL de TE/LiAc par 0,1  $A^{600}$ .
- Ajouter l'ADN de sperme de saumon (2 µL pour 20 µL de cellules).
- Répartir les levures dans une plaque 96 puits, 20 µL par puits.
- Ajouter 8 µL de vecteur d'intérêt et 100 µL de la solution TE/LiAc/PEG à chaque puit.
- Incuber à 30°C pendant 30 minutes puis à 42°C pendant 15 minutes.

- Centrifuger 5 minutes à 1800 rpm.
- Retirer le surnageant. Ajouter 110 µL d'eau stérile et retirer immédiatement 100 µL.
- Resuspendre les cellules dans les 10 µL restants et mettre environ 5 µL sur milieu sélectif approprié.

#### Solution de TE/LiAc

- 1/10 du volume total : LiAc 1 M
- 1/10 du volume total : 10x TE (100 mM Tris, 10 mM EDTA)
- 8/10 du volume total : ddH<sub>2</sub>O

#### Solution de TE/LiAc/PEG

- 1/10 du volume total : LiAc 1M
- 1/10 du volume total : 10x TE (100 mM Tris-HCl pH8.0, 10 mM EDTA)
- 8/10 du volume total : 50% PEG (polyéthylèneglycol, MW 3350, Sigma Catalog No. P3640)

## **6. Crible double-hybride (2H) :**

### **6.1. Principe**

Le système double-hybride permet d'identifier une interaction protéine-protéine *in vivo* par reconstitution d'un activateur de l'activité transcriptionnelle. La méthode, basée sur les propriétés de la protéine Gal4 trouvée chez la levure *Saccharomyces Cerevisiae*, consiste en la séparation du domaine responsable de la liaison à l'ADN et du domaine de transactivation.

La protéine associée au domaine de liaison à l'ADN de Gal4 est appelée "appât". L'appât est généralement une protéine ou un fragment de protéine dont on recherche les partenaires cellulaires. La protéine associée au domaine activateur de la transcription de Gal4 est appelée "proie". La proie est souvent inconnue (Figure 34).

### **6.2. Le système 2H par co-transformation**

Les levures utilisées ont auparavant été transformées avec le vecteur DB-X codant une protéine de fusion où le domaine de liaison à l'ADN de GAL4 est en phase avec une des ORFs "full length". Ces levures vont subir une seconde transformation avec la banque de fragments et être étalées sur des plaques de sélection 3AT (plaques SC –leu –his –trp + 20 mM de 3AT).

Les quantités fournies dans ce protocole correspondent à un crible de 5 plaques 3AT.

**Jour 1 :** Faire croître les levures "overnight" à 30°C, sous agitation (250 rpm) dans 100 mL de milieu YEPD liquide. Le lendemain matin, l'absorbance à 600 nm ( $A^{600}$ ) sera entre 0,1 et 0,5.

**Jour 2 :**

- Mesurer l' $A^{600}$ .
- Préparer l'ADN de sperme de saumon, la solution de TE/LiAc et la solution de TE/LiAc/PEG
- Verser 50 mL de levures dans un Falcon tube de 50 mL (Falcon Cat. No. 352098).
- Centrifuger 3 minutes à 1800 rpm puis resuspendre avec 50 mL d'eau stérile.
- Centrifuger 3 minutes à 1800 rpm puis resuspendre avec 10 mL de TE/LiAc.
- Centrifuger 3 minutes à 1800 rpm puis resuspendre dans un volume total de 0,2 mL de TE/LiAc par 0,1  $A^{600}$ .
- Remplir un tube de micro-centrifugation de 2 mL (USA Scientific Catalog No. 1420-2700) avec :
  - 266 µL de cellules
  - 26,6 µL d'ADN de sperme de saumon
  - 5 µg d'ADN provenant du "fragmentome 44" dilué ou non
  - 1200 µL de la solution TE/LiAc/PEG
- Vortexer quelques secondes.
- Incuber à 30°C pendant 30 minutes puis à 42°C pendant 15 minutes.

- Centrifuger pendant 3 minutes à 1800 rpm.
- Resuspension dans 1500  $\mu$ L d'eau stérile.
- Étaler 300  $\mu$ L de la solution finale par plaque 3AT
- Contrôle de co-transformation : ajouter 10  $\mu$ L de la solution finale à 10 mL d'eau stérile et étaler 300  $\mu$ L de ce mélange sur plaques –leu –trp (efficacité = nr. de colonies X 30 000).

**Jour 4 :** Compter le nombre de colonies capables de croître sur les plaques contrôle.

**Jour 7 :** Prendre avec un tips quelques cellules des colonies positives des plaques 3AT et les suspendre dans 50  $\mu$ L d'eau stérile. Mettre ensuite 5  $\mu$ L de ce mélange sur de nouvelles plaques 3AT. Exécuter une réplique velours nettoyante pour ne laisser que quelques cellules par spots, ce qui facilite l'observation d'une croissance éventuelle plus facile à visualiser). Pour ce faire, presser la plaque sur un velours autoclavé.

**Jour 10 :** Réplique velours sur de nouvelles plaques 3AT

**Jour 13 :** Réplique velours sur de nouvelles plaques 3AT

**Jour 16 :** Réplique velours sur les plaques suivantes :

- –leu –trp (comme sauvegarde)
- YEPD sur un filtre de nitrocellulose (pour le test  $\beta$ -gal)
- YEPD (pour les réactions PCR avant le séquençage)
- 3AT (pour le test de croissance sur 3AT)

Ajouter à chaque fois cinq contrôles au bas des plaques (Figure 35)

**Jour 17 :** Test  $\beta$ -gal sur le filtre de nitrocellulose pour visualiser les clones présentant une interaction.

**Jour 18 :**

- Lyser les clones de la plaque YEPD, amplifier par PCR l'insert du vecteur AD-Y et séquencer.
- Noter le phénotype du test de croissance sur 3AT.
- Mettre la plaque –leu –trp à 4°C pour une utilisation future.

La Table 12 résume la fonction des différentes plaques utilisées pendant un crible 2H.

### 6.3. Phénotypage des colonies :

Le phénotypage permet de visualiser les colonies qui présentent une interaction entre les protéines de fusion AD-Y et BD-X (test  $\beta$ -gal ou test de croissance sur 3AT ; Figure 34) ou de détecter d'éventuels faux positifs (test d'auto-activation ; Figure 36). L'utilisation de deux gènes rapporteurs différents permet de discriminer certains faux positifs.

#### 6.3.1. Test $\beta$ -gal :

1. Une solution stock de tampon Z est préparée en mélangeant :

- 16.1 g Na<sub>2</sub>HPO<sub>4</sub>·7H<sub>2</sub>O
- 5.5 g NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O
- 0.75 g KCl
- 0.246 g MgSO<sub>4</sub>·7H<sub>2</sub>O

Dissoudre dans 1L d'eau distillée et ajuster le pH à 7.0.

2. Répliquer à l'aide d'un velours les levures de la plaque –leu –trp sur un filtre de nitrocellulose (Osmonics Catalog No. WP4HY13750) placé à la surface d'une plaque YEPD. Incuber pendant 18 à 24h à 30°C.
3. Pour chaque plaque testée, mettre deux filtres de papier Whatman (Whatman Catalog No. 1454125) dans une boîte de Pétri vide de 15 cm.
4. Verser environ 200 mL d'azote liquide dans un bac à glace, mettre sous hotte.
5. Transférer le tampon Z dans les boîtes de Pétri contenant les papiers Whatman, s'assurer que l'entière du papier est mouillée et veiller à éliminer toutes les bulles d'air avec une pince à épiler. Pour chaque boîte, utiliser 6 mL de tampon Z auquel ont été ajoutés 11  $\mu$ L de  $\beta$ -mercaptoéthanol et 100  $\mu$ L de X-gal 4% [10 mg d'X-gal (5-bromo-5-chloro-3-indolyl- $\beta$ -Dgalactoside) dans 100  $\mu$ L de N,N-diméthyl formamide]

6. Avec la pince à épiler, saisir le filtre de nitrocellulose qui porte les colonies de levures et le plonger pendant 10 secondes dans l'azote liquide.
7. Mettre prudemment le filtre sur les papiers Whatman et retirer toutes les bulles d'air.
8. Incuber à 37°C, vérifier régulièrement et si nécessaire prendre des photos.
9. Incuber à 37°C "overnight" puis lire le résultat.

Les interacteurs forts sont colorés en bleu après seulement une heure (comme par exemple le contrôle 3, 4 et 5). Les plus faibles interacteurs ne sont visibles que après 24h (contrôle 2).

### **6.3.2. Test de croissance sur 3AT**

Le 3-amino-triazole (3AT), ajouté à une plaque –leu –trp –his, inhibe par compétition les produits du gène HIS3 et augmente ainsi le sévérité de la sélection. Seules les levures où il y a une interaction entre les deux protéines de fusion et donc une transcription active du gène HIS3 peuvent croître sur ce type de milieu. Ce test peut être utilisé selon deux approches : un crible et un test de phénotypage. Dans le cas du crible, les levures issues de la co-transformation sont étalées sur plaques 3AT et 5 jours plus tard, les cellules capables de pousser sont récoltées et transférées sur une nouvelle plaque 3AT. Continuer la croissance sur milieu 3AT et donc la sélection de levures montrant une interaction pendant une dizaine de jours. (Vidalain et al. 2003). Pour le test de phénotypage, les levures sont répliquées avec un velours sur une nouvelle plaque 3AT. Une réplique velours nettoyante (presser la plaque sur un velours autoclavé) permet de ne laisser que quelques cellules par spots et donc faciliter la visualisation d'une croissance. Les résultats d'interaction sont lus après une incubation à 30°C de 48h.

### **6.3.3. Test d'auto-activation**

Certaines protéines de fusion BD-X peuvent posséder une activité intrinsèque de transactivation. Ces d'auto-activateurs peuvent facilement être éliminés en début de crible en testant l'activité des gènes reporteurs (HIS3 et LacZ) dans des cellules ne contenant que le plasmide BD-X (Figure 36).

## **6.4. PCR sur les colonies de levures :**

1. Resuspendre quelques cellules dans 12 µL de tampon de lyse
2. Incuber à 37°C pendant 5 minutes puis à 95°C pendant 5 nouvelles minutes.
3. Diluer l'ADN 5 à 10 fois avec de l'eau et utiliser 3 µL dans une réaction PCR d'un volume final de 50 µL

### **Tampon de lyse :**

- 0.1 M NaPO<sub>4</sub> tampon de pH 7.4
- 2.5 mg/mL Zymolase 20T (21100 U/g, Seikagaku Corporation)

Remarque : il est important d'utiliser des levures bien "fraîches". Pour ce faire, faire croître les levures "overnight" à une température de 30°C sur des plaques de milieu complet YEPD.

## 7. Souches utilisées

Les souches de levures et de bactéries utilisées sont répertoriées à la Table 13.

## 8. Réactions de MultiSite Gateway™

Elle s'inspire du processus naturel d'intégration/excision du phage lambda dans génome d'*E. coli* (Figure 37). Les sites naturels de recombinaisons attB, L, R et P ont été modifiés de façon à augmenter l'efficacité de recombinaison (proche de 100%) et à permettre le clonage directionnel (réaction de recombinaison entre paires de séquences partenaires). La première étape consiste à obtenir par PCR l'ORF à cloner avec, à chacune de ses extrémités, deux sites de recombinaison attB : attB1 et attB2. Ces deux sites sont légèrement différents l'un de l'autre pour assurer tout au long du processus un clonage directionnel.

### 8.1. Réaction BP Clonage

Il s'agit de cloner le produit PCR obtenu par une première réaction de recombinaison : la réaction BP entre les produits PCR et le vecteur pDONr233. On obtient alors un "clone d'entrée" dans lequel l'ORF est maintenant flanquée de deux sites attL : attL1 et attL2.

Réactifs	Volume en µL
Tampon BP 5x	2
pDONr233	1
Produit PCR	2
TE 1x 10 mM, pH8.0	3
BP Clonase	2

Ne pas laisser la BP Clonase au à température ambiante trop longtemps. Laisser la réaction se réaliser "overnight" puis transformer des bactéries avec ce mélange (Figure 38).

### 8.2. Réaction LR Clonage

La recombinaison Gateway™ LR transfère l'ORF du clone d'entrée (vecteur pDONr233) vers le vecteur d'expression 2H pDEST™32 générant une fusion en phase avec le domaine de liaison à l'ADN de Gal4. (Le vecteur d'expression est muni de sites attR1 et attR2 qui recombinent spécifiquement avec les sites attL1 et attL2 du "clone d'entrée").

Réactifs	Volume en µL
Tampon LR 5x	2
Vecteur d'entrée	3
Vecteur de destination	2
LR Clonase	2

Ne pas laisser la BP Clonase au à température ambiante trop longtemps. Laisser la réaction se réaliser "overnight" puis transformer des bactéries avec ce mélange (Figure 38).

**Bibliographie**

- (1) Bhardwaj, A. & Wilkinson, M.F. A metabolic enzyme doing double duty as a transcription factor. *Bioessays* **27**, 467-71 (2005).
- (2) Bissell, M.J. & Radisky, D. Putting tumours in context. *Nat Rev Cancer* **1**, 46-54 (2001).
- (3) Copley, R.R., Doerks, T., Letunic, I. & Bork, P. Protein domain analysis in the era of complete genomes. *FEBS Lett* **513**, 129-34 (2002).
- (4) Fernandez, A.G. et al. New genes with roles in the *C. elegans* embryo revealed using RNAi of ovary-enriched ORFeome clones. *Genome Res* **15**, 250-9 (2005).
- (5) Fraser, H.B. Modularity and evolutionary constraint on proteins. *Nat Genet* **37**, 351-2 (2005).
- (6) Fromont-Racine, M., Rain, J.C. & Legrain, P. Building protein-protein networks by two-hybrid mating strategy. *Methods Enzymol* **350**, 513-24 (2002).
- (7) Hall, D.A. et al. Regulation of gene expression by a metabolic enzyme. *Science* **306**, 482-4 (2004).
- (8) Hanahan, D. Studies on transformation of *Escherichia coli* with plasmids. *J Mol Biol* **166**, 557-80 (1983).
- (9) Hartwell, L.H., Hopfield, J.J., Leibler, S. & Murray, A.W. From molecular to modular cell biology. *Nature* **402**, C47-52 (1999).
- (10) Kamath, R.S. et al. Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* **421**, 231-7 (2003).
- (11) Lamesch, P. et al. *C. elegans* ORFeome version 3.1: increasing the coverage of ORFeome resources with improved gene predictions. *Genome Res* **14**, 2064-9 (2004).
- (12) Li, S. et al. A map of the interactome network of the metazoan *C. elegans*. *Science* **303**, 540-3 (2004).



- 
- (13) Li, X. & Burgers, P.M. Cloning and characterization of the essential *Saccharomyces cerevisiae* RFC4 gene encoding the 37-kDa subunit of replication factor C. *J Biol Chem* **269**, 21880-4 (1994).
  - (14) Lyczak, R., Gomes, J.E. & Bowerman, B. Heads or tails: cell polarity and axis formation in the early *Caenorhabditis elegans* embryo. *Dev Cell* **3**, 157-66 (2002).
  - (15) Misteli, T. The concept of self-organization in cellular architecture. *J Cell Biol* **155**, 181-5 (2001).
  - (16) Oltvai, Z.N. & Barabasi, A.L. Systems biology. Life's complexity pyramid. *Science* **298**, 763-4 (2002).
  - (17) Pawson, T. & Nash, P. Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-52 (2003).
  - (18) Pazos, F. & Valencia, A. In silico two-hybrid system for the selection of physically interacting protein pairs. *Proteins* **47**, 219-27 (2002).
  - (19) Redon, C. et al. Histone H2A variants H2AX and H2AZ. *Curr Opin Genet Dev* **12**, 162-9 (2002).
  - (20) Reinke, V. et al. A global profile of germline gene expression in *C. elegans*. *Mol Cell* **6**, 605-16 (2000).
  - (21) Vallee, R.B., Tai, C. & Faulkner, N.E. LIS1: cellular function of a disease-causing gene. *Trends Cell Biol* **11**, 155-60 (2001).
  - (22) Venclovas, C., Colvin, M.E. & Thelen, M.P. Molecular modeling-based analysis of interactions in the RFC-dependent clamp-loading process. *Protein Sci* **11**, 2403-16 (2002).
  - (23) Vidalain, P.O., Boxem, M., Ge, H., Li, S. & Vidal, M. Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* **32**, 363-70 (2004).
  - (24) Walhout, A.J. et al. Integrating interactome, phenome, and transcriptome mapping data for the *C. elegans* germline. *Curr Biol* **12**, 1952-8 (2002).